THE COMPACTNESS OF THE RIEMANN MANIFOLD OF AN ABSTRACT FIELD OF ALGEBRAIC FUNCTIONS

OSCAR ZARISKI

1. The existence of finite resolving systems. In an earlier paper[1] we have announced the result that the existence of a resolving system of the Riemann manifold of an abstract field of algebraic functions (in any number of variables) or—what is the same—the local uniformization theorem[2] implies the existence of *finite* resolving systems of the Riemann manifold. We have proved this result for algebraic surfaces by arithmetic considerations.[1] The proof for the general case of varieties, which at that time was in our possession,[3] and which we have promised to publish in a subsequent paper, was of similar nature, that is, it was based upon considerations involving the structure of certain infinite sequences of quotient rings. However, we have succeeded lately in finding a much simpler proof which is based on topological considerations.

Let Σ be a field of algebraic functions of several variables, over an arbitrary ground field k. By the Riemann manifold $\boldsymbol{M}$ of Σ we mean the totality of places of Σ, that is, the totality of zero-dimensional valuations v of Σ/k. If V is a projective model of Σ/k, and if H is any subset of V, we denote by $N(H)$ the subset of $\boldsymbol{M}$ consisting of those valuations $\mathfrak{v}$ which have center in H. By a resolving system of $\boldsymbol{M}$ we mean a collection $\mathfrak{B}=\{V_\alpha\}$ of projective models (finite or infinite in number) with the property that for any v in $\boldsymbol{M}$ there exists a V_α in $\mathfrak{B}$ such that the center of v on V_α is a simple point.

The topology which we introduce in $\boldsymbol{M}$ is simply this: *we choose as a basis for the closed sets of* $\boldsymbol{M}$ *the sets* $N(W)$, *where* W *is any algebraic subvariety of any projective model of* Σ. We prove that if topologized in this fashion, *the set* $\boldsymbol{M}$ *is a compact*[4] *topological space.* From this the result announced above follows immediately. For if $\{V_\alpha\}$ is a resolving system, and if we denote by S_α the singular locus of V_α, then $N(V_\alpha-S_\alpha)$ is an open set and $\{N(V_\alpha-S_\alpha)\}$ is an open covering of $\boldsymbol{M}$.

Received by the editors April 10, 1944.

[1] *A simplified proof for the resolution of singularities of an algebraic surface*, Ann. of Math. vol. 43 (1942) p. 583.

[2] See loc. cit. footnote 1.

[3] That proof was presented by us at a seminar in algebraic geometry at Johns Hopkins in 1942.

[4] We use the term compact in the same sense as it is used by S. Lefschetz in his *Algebraic topology* (Amer. Math. Soc. Colloquium Publications, vol. 27, 1942). The old term is bicompact.

Hence this covering contains a finite subcovering $\{N(V_i - S_i)\}$, $i = 1, 2, \cdots, m$, and this means that $\{V_1, V_2, \cdots, V_m\}$ is a finite resolving system of $\boldsymbol{M}$.

The proof of compactness of $\boldsymbol{M}$ given in the next section is based in part on some simple algebro-geometric considerations, and in part on a theorem of Steenrod[5] on the compactness of the limit space of an inverse system of compact T_1-spaces.

2. **The Riemann manifold as the limit space of an inverse system.** Let $\mathfrak{V} = \{V_\alpha\}$ be the collection of *all* projective models of Σ/k. By a point of V_α we mean a zero-dimensional prime ideal in a suitable coördinate ring of V_α, or, in other terms, a point is a prime one-dimensional *homogeneous* ideal in the ring of homogeneous coördinates of the general point of V_α. This defines V_α set-theoretically as a set of points. We topologize V_α by choosing as closed sets the algebraic subarieties of V_α. It is obvious that V_α then becomes a compact topological space in which points are closed sets (whence V_α is a T_1-space; however, it is not a Hausdorff space).

If V_α and V_β are two projective models of Σ/k, we denote by π_α^β the transformation of V_β onto V_α defined by the birational correspondence between V_α and V_β. We define a partial ordering $\prec$ of the collection $\mathfrak{V}$ as follows: $V_\alpha \prec V_\beta$ if whenever P_α and P_β are corresponding points of V_α and V_β under π_α^β then $Q(P_\alpha) \subseteq Q(P_\beta)$. Here $Q(P)$ denotes the quotient ring of P. It is clear that if $V_\alpha \prec V_\beta$ then π_α^β is a single-valued continuous and *closed* mapping. Moreover, if V_α and V_β are arbitrary projective models of Σ/k, and if V_γ denotes the join[6] of V_α and V_β, then $V_\alpha \prec V_\gamma$ and $V_\beta \prec V_\gamma$. Hence we have here an inverse system $\{V_\alpha; \pi_\alpha^\beta \mid V_\alpha \prec V_\beta\}$ of compact T_1-spaces. Let $\boldsymbol{M}$ be the limit space of the system. By Steenrod's theorem $\boldsymbol{M}$ is compact. Every point P^* of $\boldsymbol{M}$ represents an infinite collection of points $\{P_\alpha\}$, $P_\alpha \in V_\alpha$, $V_\alpha \in \mathfrak{V}$, with the property that if $V_\alpha \prec V_\beta$ then $Q(P_\alpha) \subseteq Q(P_\beta)$. We shall denote by π_α^* the mapping $P^* \to P_\alpha$ of $\boldsymbol{M}$ into V_α. If V_α is any projective model of Σ/k and if W is any algebraic subvariety of V_α, then $(\pi_\alpha^*)^{-1}W$ is a closed subset of $\boldsymbol{M}$, and the closed sets obtained in this fashion form a basis for the closed subsets of $\boldsymbol{M}$.

The compactness of the Riemann manifold of Σ/k and the implications stated in the preceding section are immediate consequences of the following theorem.

THEOREM. *There is a* (1, 1) *correspondence between the points* P^*

[5] N. E. Steenrod, *Universal homology groups*, Amer. J. Math. vol. 58 (1936) p. 666.

[6] See our paper *Foundations of a general theory of birational correspondences*, Trans. Amer. Math. Soc. vol. 53 (1943) p. 516.

of M and the zero-dimensional valuations v of the field Σ/k. If P^* and v are corresponding elements, and if V_a is any projective model of Σ/k, then $\pi_a^* P^*$ is the center of v on V_a.

PROOF. Let v be a zero-dimensional valuation of Σ/k and let $P_{a,v}$ be the center of v on any given projective model V_a of Σ/k. For any two projective models V_a, V_b it is then true that $P_{a,v}$ and $P_{b,v}$ are corresponding points in the birational correspondence π_a^b. Hence $P_v^* = \{P_{a,v}\}$ is a point of M. *Thus every zero-dimensional valuation v determines uniquely a point P_v^* of M.*

If v_1 and v_2 are two distinct zero-dimensional valuations, then there exists at least one projective model V_a such that $P_{a,v_1} \neq P_{a,v_2}$. *Hence if $v_1 \neq v_2$ than $P_{v_1}^* \neq P_{v_2}^*$.*

Now let P^* be an arbitrary point of M, $P^* = \{P_a\}$. We denote by $\mathfrak{B}$ the least ring containing the quotient rings $Q(P_a)$. Let V_b be a fixed projective model of Σ/k and let $P_b = \pi_b^* P^*$. *We assert that if ω is a non-unit in $Q(P_b)$ then* $1/\omega \notin \mathfrak{B}$. For assume that $1/\omega \in \mathfrak{B}$. Then $1/\omega$ will belong to the ring generated by a finite number of quotient rings $Q(P_a)$, say $Q(P_{a_1})$, $Q(P_{a_2})$, $\cdots$, $Q(P_{a_m})$. Let V_c be the join of the varieties V_b, V_{a_1}, V_{a_2}, $\cdots$, V_{a_m} and let $P_c = \pi_c^* P^*$. Since $\pi_{a_i}^* P^* = P_{a_i}$ and $\pi_b^* P^* = P_b$, we have $\pi_{a_i}^c P_c = P_{a_i}$ and $\pi_b^c P_c = P_b$, and hence $Q(P_{a_i}) \subseteq Q(P_c)$, $Q(P_b) \subseteq Q(P_c)$. Therefore $1/\omega \in Q(P_c)$. This is a contradiction since any non-unit of $Q(P_b)$ is obviously also a non-unit in $Q(P_c)$.

We have therefore shown that $\mathfrak{B}$ is a proper ring (not a field). We now show that $\mathfrak{B}$ *is a valuation ring.* For this it is sufficient to show that if ξ is any element of Σ then either $\xi \in \mathfrak{B}$ or $1/\xi \in \mathfrak{B}$. We consider again a fixed projective model V_b of Σ/k. We select a system of nonhomogeneous coördinates $x_1, x_2, \cdots, x_n$ of the general point of V_b in such a fashion that the point P_b $(=\pi_b^* P^*)$ is at finite distance with respect to these coördinates. Let V_d be the projective model whose general point has as nonhomogeneous coördinates the elements $x_1, x_2, \cdots, x_n, \xi$. If the point P_d $(=\pi_d^* P^*)$ is at finite distance with respect to these coördinates, then $\xi \in Q(P_d) \subseteq \mathfrak{B}$. If P_d is a point at infinity, we observe first of all that in the above proof of our assertion $1/\omega \notin \mathfrak{B}$ we have shown incidentally the following: if V_a and V_b are *any* two projective models of Σ/k and if $\pi_a^* P^* = P_a$ and $\pi_b^* P^* = P_b$, then P_a and P_b are *corresponding* points of the birational correspondence between V_a and V_b. For on the join V_c of V_a and V_b we have the point $P_c = \pi_c^* P^*$ and the relations $Q(P_c) \supseteq Q(P_a)$, $Q(P_c) \supseteq Q(P_b)$. These relations show that if v is any zero-dimensional valuation whose center on V_c is the point P_c, then the center of v on V_a is P_a and its center on V_b is P_b. Hence P_a and P_b are indeed

corresponding points.[7] With this observation in mind, let v be a zero-dimensional valuation whose center on V_b is the point P_b and whose center on V_d is the point P_d. Since P_b is at finite distance, we have $v(x_i) \geqq 0$, $i=1, 2, \cdots, n$. Since P_d is at infinity, we must have $v(\xi) < 0$. Hence $v(1/\xi) > 0$, $v(x_i/\xi) > 0$, and this shows that if we take $1/\xi, x_1/\xi, \cdots, x_n/\xi$ as nonhomogeneous coördinates of the general point of V_d, then P_d is at finite distance. Hence $1/\xi \in Q(P_d) \subseteq \mathfrak{B}$. This completes the proof of our assertion that $\mathfrak{B}$ is a valuation ring.

Let v be the valuation defined by the valuation ring $\mathfrak{B}$. We assert that *v is zero-dimensional.* For let v be of dimension s. We can find a projective model V_b on which the center of v is an s-dimensional variety W. If $P_b = \pi_b^* P^*$, then $Q(P_b) \subseteq \mathfrak{B}$ and this implies that $P_b \in W$.[8] If $s > 0$, then we can find a non-unit ω in $Q(P_b)$ such that $\omega \neq 0$ on W, whence $1/\omega \in Q(W) \subseteq \mathfrak{B}$, a contradiction. Hence $s = 0$, as asserted.

The above relation $P_b \in W$ implies now $P_b = W$. This is true for any projective model V_b, that is, the center of v on any projective model V_b is the point $P_b = \pi_b^* P^*$. This completes the proof of the theorem.

3. **A generalization. Infinite direct products of projective lines.** The idea of topologizing an algebraic variety V by choosing as closed sets the algebraic subvarieties of V can be used with good effect in order to topologize the set M^* of all homomorphic mappings of any abstract field A into another abstract field K. In this general case we are dealing essentially with a generalization of the concept of the Riemann manifold of a field of algebraic functions (see the Remark at the end of the paper). We begin with some topological preliminaries.

Let $\{R_a\}$ be a system of compact topological spaces indexed by a set $A = \{a\}$. We assume that each R_a is a T_1-space; that is, that the points of R_a are closed sets. Elements of A shall be denoted by small Latin letters, $a, b, c, \cdots$; subsets of A shall be denoted by small Greek letters, $\alpha, \beta, \gamma, \cdots$. If α is a subset of A we shall denote by R_α the direct product $P_{a \in \alpha} R_a$. If $\alpha \subset \beta$ we denote by π_α^β *the projection* of R_β onto R_α. Finally, elements of R_α and R_α shall be denoted by $x_a, y_a, z_a, \cdots$ and by $x_\alpha, y_\alpha, z_\alpha, \cdots$ respectively. If $a \in \alpha$ and if $\pi_a^\alpha x_\alpha = x_a$, then x_a shall be referred to as the a-component of x_α.

We assume that for each *finite* subset α of A a topology has been assigned to R_α and that the following three conditions are satisfied: (1) *R_α is a compact topological space*; (2) *if $\alpha \subset \beta$ then π_α^β is a closed*

[7] See our definition of corresponding points of a birational transformation, loc. cit. footnote 6, p. 505.

[8] See loc. cit. footnote 6, Theorem 3, p. 497.

mapping (mapping = single-valued continuous transformation); (3) *if α is a set with one element a then the topology assigned to R_α is exactly the topology of R_a.* It is clear that in virtue of these two conditions R_α is a T_1-space. For if x_a is the a-component of x_α, then $(\pi_a^\alpha)^{-1}x_a$ is closed and x_α is the intersection of the closed sets $(\pi_a^\alpha)^{-1}x_a$, $a \in \alpha$.

If we consider only finite subsets α of A and if we define a partial ordering in the collection $\{R_\alpha\}$ by setting $R_\alpha < R_\beta$ if $\alpha \subset \beta$, then we have an inverse system $\{R_\alpha; \pi_\alpha^\beta\}$. It is clear that set-theoretically the limit space R^* of the system coincides with the direct product $R^* = P_{a \in A} R_a$. However, the topology in R^* is not necessarily the usual topology of the product space, for our topology in R^* depends not only on the topology of each factor R_a but also on the topology which has been assigned to each R_α, for α any finite subset of A.

Our space R^* is compact, by Steenrod's theorem. We are dealing here with a special case of Steenrod's theorem, and the proof of the compactness of R^* can be somewhat simplified. For this reason, and also for the convenience of the reader, we shall include here a proof of the compactness of R^*.

We have to show that if a family of closed sets in R^* has the finite intersection property (that is, if every finite subfamily has a non-empty intersection), then the intersection of the entire family is non-empty. It will be sufficient to prove this for families of basic closed sets F_α^*, $F_\alpha^* = \pi_\alpha^{-1} F_\alpha$, where F_α denotes a closed set in R_α and where π_α is the projection of R^* onto R_α. Let then $\{F_\alpha^*\}$ be a family $\mathfrak{F}$ of basic closed sets which has the finite intersection property. By Zorn's lemma the family $\{F_\alpha^*\}$ is contained in a maximal family $\{G_\alpha^*\}$ of basic closed sets which has the finite intersection property. It will be sufficient to show that $\bigcap G_\alpha^*$ is non-empty. We shall therefore assume that our original family $\{F_\alpha^*\}$ is not contained properly in another family of basic closed sets which has the finite intersection property.[9]

We first observe that *the intersection of any finite collection of basic closed sets is itself a basic closed set.* For let $\{\pi_{\alpha_i}^{-1} F_{\alpha_i}\}$ be a finite collection of basic closed sets. We put $\alpha = \bigcup \alpha_i$, $F_\alpha = \bigcap (\pi_{\alpha_i}^\alpha)^{-1} F_{\alpha_i}$. Then it is clear that $\bigcap \pi_{\alpha_i}^{-1} F_{\alpha_i} = \pi_\alpha^{-1} F_\alpha$.

In virtue of this remark and in virtue of the maximality property

[9] The idea of passing to a maximal family is taken from the proof of Tychonoff's theorem as given in Lefschetz, *Algebraic topology*, p. 19. There is only this difference: the maximal family in Lefschetz is not a family of closed sets, while ours is. This modification of the proof succeeds because we restrict ourselves to families of basic closed sets and because in our case the mappings π_α^β are closed.

of the given family $\mathfrak{F}$, it follows that every finite intersection of sets in $\mathfrak{F}$ is again in the family $\mathfrak{F}$.

For any element a in A and for any member F_α^* in $\mathfrak{F}$ let $\pi_a F_\alpha^* = F_{\alpha,a}$. If $a \notin \alpha$ then it is clear that $F_{\alpha,a} = R_a$, for then the a-component of the points of F_α^* is not restricted. If $a \in \alpha$ and if $F_\alpha^* = \pi_a^{-1} F_a$, then $F_{\alpha,a} = \pi_a^\alpha F_a$. In either case $F_{\alpha,a}$ is a closed set in R_a, for we have assumed that π_a^β is closed whenever $\alpha \subset \beta$. For a given a the family $\mathfrak{F}_a$ of closed sets $\{F_{\alpha,a}\}$ has the finite intersection property. Since R_a is compact, the intersection $\bigcap_\alpha F_{\alpha,a}$ is non-empty. Let x_a be a point common to all the sets in $\mathfrak{F}_a$. Then $\pi_a^{-1} x_a$ is a basic closed set (since R_a is a T_1-space) which meets every set F_α^* in $\mathfrak{F}$. Consequently $\pi_a^{-1} x_a \in \mathfrak{F}$, $x_a \in \mathfrak{F}_a$, and the intersection $\bigcap_\alpha F_{\alpha,a}$ consists only of the point x_a.

Let then $x = \{x_a\}$, where $x_a = \bigcap_\alpha F_{\alpha,a}$. We show that x is a common point of the sets F_α^* in $\mathfrak{F}$. Since $\pi_a^{-1} x_a \in \mathfrak{F}$, for any a, it follows that $\bigcap_{a \in \alpha} \pi_a^{-1} x_a \in \mathfrak{F}$, that is, $\pi_\alpha^{-1} x_\alpha \in \mathfrak{F}$, where $x_\alpha = \pi_\alpha x$. Therefore $\pi_\alpha^{-1} x_\alpha$ meets F_α^*, that is, $\pi_\alpha^{-1} F_\alpha$; hence $x_\alpha \in F_\alpha$ and $x \in \pi_\alpha^{-1} x_\alpha \subset \pi_\alpha^{-1} F_\alpha = F_\alpha^*$, q.e.d.

Now let K be a fixed abstract field and let the sets R_a be projective lines over K, so that the points of each set R_a are in (1, 1) correspondence with the elements of K together with the symbol ∞. We topologize R_a by choosing as closed sets the finite subsets of R_a. Then each R_a becomes a compact topological T_1-space.

We still have to topologize each set R_α, for α a finite subset of A. For this purpose we introduce on each line R_a a pair of homogeneous coördinates x_{a1}, x_{a2} and we define an algebraic variety V_α by the following parametric equations (in which the $X_{(\delta)}^{(\alpha)}$ denote the homogeneous coördinates of the general point of V_α):

$$\rho \cdot X_{\epsilon_1 \epsilon_2 \cdots \epsilon_n}^{(\alpha)} = x_{a_1 \epsilon_1} x_{a_2 \epsilon_2} \cdots x_{a_n \epsilon_n}, \tag{1}$$

where $\alpha = \{a_1, a_2, \cdots, a_r\}$ and where each ϵ_j can take the values 1 or 2. It is well known that V_α is a Segre variety, of dimension n, immersed in a projective space of dimension $2^n - 1$. The points of V_α are in (1, 1) correspondence with n-tuples of ratios $\{x_{a2}/x_{a1}\}$, $a \in \alpha$, that is, with the points of the direct product $R_\alpha = R_{a_1} \times R_{a_2} \times \cdots \times R_{a_n}$. It should be noted that here we only consider points X^α whose homogeneous coördinates are in K. *We topologize V_α by choosing as closed sets the algebraic subvarieties of V_α.* Then it is clear that each V_α becomes a compact topological T_1-space.

If $\alpha = \{a_1, a_2, \cdots, a_n\}$ and if β is a subset of α, say if $\beta = \{a_1, a_2, \cdots, a_m\}$, $m < n$, then the projection π_β^α of V_α onto V_β is given by the equations:

$$X_{\epsilon_1 \epsilon_2 \cdots \epsilon_m}^{(\beta)} : X_{\delta_1 \delta_2 \cdots \delta_m}^{(\beta)} = X_{\epsilon_1 \epsilon_2 \cdots \epsilon_m \gamma_{m+1} \cdots \gamma_n}^{(\alpha)} : X_{\delta_1 \delta_2 \cdots \delta_m \gamma_{m+1} \cdots \gamma_n}^{(\alpha)},$$

where each ϵ, δ and γ can take *independently* the values 1 or 2. Thus π_β^α is a single-valued *rational* transformation of V_α onto V_β, and therefore π_β^α is closed and $(\pi_\beta^\alpha)^{-1}$ is open. It is clear that the closed sets in the infinite direct product R^*, as defined above, are the sets defined by (finite or infinite) systems of homogeneous equations, each equation involving the variables $X^{(\alpha)}$ relative to some finite subset α of A.

4. The space of homomorphic mappings of one abstract field into another. We now further specialize our application by assuming that the set A is a field. The space R^* is then the space of all single-valued transformations $x^*: a \to x_a = x_{a1}/x_{a2}$, of the field A into the set consisting of the elements of the field K and of the symbol ∞. We shall now express in an appropriate homogeneous form the conditions that a given mapping x^* be a homomorphism. Let α be a subset of A consisting of three elements, $\alpha = \{a_1, a_2, a_3\}$. On the corresponding variety V_α let F_{a_1,a_2,a_3} be the algebraic subvariety obtained by imposing on the 6 parameters x_{i1}, x_{i2}, $i = a_1, a_2, a_3$, the following condition:

$$x_{a_1 1}x_{a_2 2}x_{a_3 2} + x_{a_1 2}x_{a_2 1}x_{a_3 2} = x_{a_1 2}x_{a_2 2}x_{a_3 1}. \tag{2}$$

Similarly we define another algebraic subvariety G_{a_1,a_2,a_3} of V_α by the equation

$$x_{a_1 1}x_{a_2 1}x_{a_3 2} = x_{a_1 2}x_{a_2 2}x_{a_3 1}. \tag{3}$$

Let $x_{a_j 1}/x_{a_j 2} = x_j$, $j = 1, 2, 3$, where x_j may be ∞. Suppose that equation (2) holds true. Then if x_1 and x_2 are both different from ∞ we find $x_3 = x_1 + x_2$. If $x_1 = \infty$ and $x_2 \neq \infty$, then $x_{a_1 2} = 0$, $x_{a_1 1} \cdot x_{a_2 2} \neq 0$, whence (2) yields $x_{a_3 2} = 0$, that is, $x_3 = \infty$. Assume now that equation (3) holds. Again we find that $x_3 = x_1 x_2$, if both x_1 and x_2 are different from ∞. If $x_1 = \infty$ and $x_2 \neq 0$, then $x_{a_1 2} = 0$, $x_{a_1 1} \cdot x_{a_2 1} \neq 0$, and (3) yields $x_{a_3 2} = 0$, that is, $x_3 = \infty$. Thus the equations (2) and (3) are the homogeneous counterparts of the equations $x_3 = x_1 + x_2$ and $x_3 = x_1 x_2$ respectively, and they include the conventions which are usually made for the symbol ∞. We can therefore assert that x^* represents a homomorphic mapping of A into (K, ∞) if and only if the following conditions are satisfied: for any three elements a_1, a_2, a_3 of A such that respectively $a_3 = a_1 + a_2$ or $a_3 = a_1 a_2$, the projection $\pi_\alpha^{-1} x^*$ (where $\alpha = \{a_1, a_2, a_3\}$) must lie respectively on F_{a_1,a_2,a_3} or on G_{a_1,a_2,a_3}. Therefore, if we denote by M the set of all *homomorphic* mappings of A into (K, ∞), we see that

$$M = \bigcap_\alpha \pi_\alpha^{-1} F_{a_1 a_2 a_3} \bigcap_\beta \pi_\beta^{-1} G_{b_1 b_2 b_3},$$

where the index α ranges over all sets $\alpha=\{a_1, a_2, a_3\}$ such that $a_3=a_1+a_2$, and the index β ranges over the sets $\beta=\{b_1, b_2, b_3\}$ such that $b_3=b_1b_2$. We see thus that M is an intersection of basic closed subsets of R^*. Hence M is closed, and since R^* is compact M is also compact.

The case which is of special interest to us is that in which K is a subfield of A. In this case we are interested in the relative homomorphisms of A into (K, ∞), that is, in the homomorphisms x^* which leave each element of K invariant. If M^* is the set of all these relative homomorphisms, then it is clear that M^* is the intersection of M with the closed set $\bigcap_{a\in K}\pi_a^{-1}a$. Here, according to our notations, $\pi_a^{-1}a$ denotes that subset of R^* which consists of the points x^* whose a-component x_a is a itself ($a\in K$). Hence also M^* is a compact space.

It is convenient to describe in algebro-geometric terms the relative topology induced in M^* by the topology of M. Let $x_1, x_2, \cdots, x_n$ be a finite set of elements of A. For each x_i we introduce a pair of homogeneous parameters x_{i1}, x_{i2} such that $x_{i1}/x_{i2}=x_i$. We consider the algebraic variety Z over K whose general point has as homogeneous coördinates the quantities $X_{(\epsilon)}$ defined by the parametric equations

$$\rho X_{\epsilon_1\epsilon_2\cdots\epsilon_n} = x_{1\epsilon_1}x_{2\epsilon_2}\cdots x_{n\epsilon_n}, \tag{4}$$

where each ϵ_j can take the values 1 or 2. If the quantities x_i are algebraically independent, then the variety Z coincides with the variety V_α defined by the equations (1), α being the subset $\{x_1, x_2, \cdots, x_n\}$ of A. But in general Z is a subvariety of V_α. If $x^*\in M^*$, then the mapping x^* of A into (K, ∞) must preserve all the algebraic relations between $x_1, x_2, \cdots, x_n$ over K, since x^* is a homomorphism. It follows that the point $\pi_\alpha x^*$ of V_α must lie on Z. Now we observe that the homomorphism x^* defines a unique valuation of A/K whose residue field is K itself and whose center on Z is the point $\pi_\alpha x^*$. Conversely, every valuation of A/K whose residue field is K defines a relative homomorphic mapping of A/K onto (K, ∞). We conclude that *if W is any algebraic subvariety of Z, then the set of all valuations of A/K having K as residue field and having center on W is a closed subset of M^**. By taking different finite subsets $\{x_1, x_2, \cdots, x_n\}$ of A and different subvarieties W of Z we obtain a family of closed subsets of M^* which form a basis for the closed subsets of M^*.

REMARK. Suppose that A is a field of algebraic functions in any number of variables, over a given ground field k. We identify the field K with the algebraically closed field determined by k. The Riemann manifold M of A is the set of all zero-dimensional valuations v

of A. By the ground field extension $k \rightarrow K$ we can embed A in a field $A' = KA$. Every relative homomorphic mapping of A' onto (K, ∞) determines uniquely a zero-dimensional valuation of A'/K, and vice versa. Every zero-dimensional valuation of A'/K induces a unique zero-dimensional valuation of A/K, but a given zero-dimensional valuation of A/K may be extendable in more than one way to a zero-dimensional valuation of A'/K. It follows that the Riemann manifold M' of A'/K coincides with the space M^* of all relative homomorphic mappings of A' onto (K, ∞) and is therefore a compact space. The Riemann manifold M of A/K is obtainable from M' by topological identification and therefore can also be converted into a compact topological space. That is precisely what we have proved in §2.

THE JOHNS HOPKINS UNIVERSITY

THE COMPLETION OF A THEOREM OF KANTOR

GERALD B. HUFF

The nature of the problem. Let a homaloidal net be defined by its order n and its multiplicities $s_1, s_2, \cdots, s_\rho$ at set P_ρ^2 of ρ general points in the plane. The positive integers $n, s_1, s_2, \cdots, s_\rho$ satisfy the equations

$$\begin{aligned} s_1^2 + s_2^2 + \cdots + s_\rho^2 - n^2 &= -1, \\ s_1 + s_2 + \cdots + s_\rho - 3n &= -3. \end{aligned} \tag{1}$$

A planar Cremona transformation C is set up by putting this net into projective correspondence with a net of lines in another plane. If a complete and regular linear system $\Sigma_{p,d}$ of dimension d, the generic curve of the system having the genus p, has the order x_0 and the multiplicities $x_1, x_2, \cdots, x_\rho$, then $x = \{x_0; x_1, \cdots, x_\rho\}$ is called the characteristic of $\Sigma_{p,d}$. The image of $\Sigma_{p,d}$ under C is another linear system of the same p, d and a characteristic $x' = \{x_0'; x_1', \cdots, x_\rho'\}$ at the set Q_ρ^2 of the fundamental points of C^{-1}. x' is related to x by the substitution

$$\begin{aligned} x_0' &= nx_0 - r_1x_1 - \cdots - r_\rho x_\rho, \\ x_1' &= s_1x_0 - \alpha_{11}x_1 - \cdots - \alpha_{1\rho}x_\rho, \\ &\cdots\cdots\cdots\cdots\cdots, \\ x_\rho' &= s_\rho x_0 - \alpha_{\rho 1}x_1 - \cdots - \alpha_{\rho\rho}x_\rho. \end{aligned} \tag{2}$$

The sets of numbers $\{s_i, \alpha_{i1}, \cdots, \alpha_{i\rho}\}$ are the characteristics of the principal curves of C at P_ρ^2 and satisfy the equations

$$\begin{aligned} \alpha_{i1}^2 + \alpha_{i2}^2 + \cdots + \alpha_{i\rho}^2 - s_i^2 &= 1, \\ \alpha_{i1} + \alpha_{i2} + \cdots + \alpha_{i\rho} - 3s_i &= -1. \end{aligned} \tag{3}$$

The linear substitution (2) has as absolute invariants the forms

$$\begin{aligned} (xx) &\equiv x_1^2 + x_2^2 + \cdots + x_\rho^2 - x_0^2; \\ (lx) &\equiv x_1 + x_2 + \cdots + x_\rho - 3x_0. \end{aligned} \tag{4}$$

The problems considered in this paper arise from the fact that the converses of two of the above statements do not hold. There are sets of positive integers satisfying equations (1) which are not associated with any homaloidal net and there are linear substitutions (2) which

Presented to the Society, April 29, 1944; received by the editors April 17, 1944.

leave (xx), (lx) invariant which are not associated with any planar Cremona transformation. A solution of (1) or a linear substitution (2) which does have the geometric meaning described above will be said to be proper.

In his prize memoir [8][1] of 1884, S. Kantor stated the following proposition: *a linear substitution* (2) *leaving* (xx) *and* (lx) *invariant and in which* n, s_i, r_j, α_{ij} *are non-negative integers is proper.* He gave two "proofs" of the assertion. In 1931, Cooolidge [4] recognized the importance of the proposition and supported it with an argument like one of Kantor's. At the time Coble [1] pointed out that the proof was not valid. In 1934, this writer [5] constructed an example ($\rho=11$) which showed that the proposition was not true; and, in 1940, by using the specific results [2] of Coble on irreducible solutions of (3), it was possible to prove [6, p. 865] that Kantor's theorem was true for $\rho<11$.

The purpose of this paper is to exhibit further necessary conditions on a proper linear substitution which will also be sufficient for all values of ρ.

1. Two lemmas.

LEMMA 1. *Let* $\{p\}=\{p_0; p_1, \cdots, p_\rho\}$ *be an integer solution of equations* (3) *such that* $p_0\geqq 0$ *and* $p_1\geqq p_2\geqq \cdots \geqq p_\rho$. *Then* $2p_0-p_1-p_2-p_3\geqq 0$.

It is easily verified that for each of $p_0=0, 1$ there is a unique solution, and that each satisfies the lemma. In the case $p_0>1$, it may be shown that $p_0>p_1$. Indeed, $p_0=p_1$ requires that $p_0=0, 1$ and it is easy to show that $p_1>p_0$ is impossible for $p_0>1$. Thus for $p_0>1$ we may write

$$p_1=p_0-b_1; \qquad p_2=p_0-b_2; \qquad p_3=p_0-b_3,$$

where $0<b_1\leqq b_2\leqq b_3$. Substitution in the quadratic relation of (3) yields

$$p_4^2+\cdots+p_\rho^2+b_1^2+b_2^2+b_3^2+2p_0^2=1+2(b_1+b_2+b_3)p_0.$$

But

$$3\leqq b_1^2+b_2^2+b_3^2+p_4^2+\cdots+p_\rho^2.$$

Adding these yields

$$2(p_0^2+1)\leqq 2(b_1+b_2+b_3)p_0.$$

[1] Numbers in brackets refer to the bibliography at the end of the paper.

Hence,

$$p_0 < b_1 + b_2 + b_3,$$

$$3p_0 - b_1 - b_2 - b_3 < 2p_0;$$

or

$$p_1 + p_2 + p_3 < 2p_0.$$

The restriction that $p_1 \geqq p_2 \geqq \cdots \geqq p_\rho$ is not essential to the proof. Indeed, it is clear that if the inequality holds for the three largest of p_i, then it must hold for any three.

LEMMA 2. *If an integer linear substitution $L(I)$ of the form (2) leaves (xx), (lx) invariant, and if the numbers $\{n; r_j\}$ constitute a proper solution of (1), then the $L(I)$ is proper.*

Since the characteristic $\{n; r_j\}$ is proper, it is associated with a Cremona transformation which has the description

$$L(C) \qquad \begin{aligned} x_0' &= nx_0 - \sigma_1 x_1 - \cdots - \sigma_\rho x_\rho, \\ x_1' &= r_1 x_1 - \beta_{11} x_1 - \cdots - \beta_{1\rho} x_\rho, \\ &\cdots\cdots\cdots\cdots\cdots\cdots, \\ x_\rho' &= r_\rho x_1 - \beta_{\rho 1} x_1 - \cdots - \beta_{\rho\rho} x_\rho. \end{aligned}$$

Consider the product[2] $L(I)L(C)$ as being in the form (2) with primed coefficients. It is a linear substitution with (xx) and (lx) as invariants and hence [3]:

(i) $\{n'; s_i'\}$ and $\{n'; r_j'\}$ satisfy (1);

(ii) $\{s_i'; \alpha_{ij}'\}$ and $\{r_j'; \alpha_{ij}'\}$ satisfy (3); and

(iii) $s_i' s_k' - \alpha_{i1}'\alpha_{k1}' - \alpha_{i2}'\alpha_{k2}' - \cdots - \alpha_{i\rho}'\alpha_{k\rho}' = 0$, $i \neq k$.

Now $n' = n^2 - r_1^2 - \cdots - r_\rho^2 = 1$, and hence, from (i), $r_1' = \cdots = r_\rho' = s_1' = \cdots = s_\rho' = 0$. Since the characteristics $\{s_i'; \alpha_{ij}'\}$ $[r_j'; \alpha_{ij}']$ satisfy (3), have $s_i' = 0$ $[r_j' = 0]$ and are integral, they must be of type $\{0; 0^{\rho-1} - 1\}$. From (iii) it is clear that if $\alpha_{ij}' = -1$, and $\alpha_{kl}' = -1$, then $j \neq l$. $L(I)L(C)$ must then be the identity or a permutation P of the letters $x_1, \cdots, x_\rho$. Then $L(I) = L^{-1}(C)$ or $L(I) = PL^{-1}(C)$. In either case, $L(I)$ is proper.

2. The proof of the theorem.

THEOREM 1. *A linear substitution L of the form (2) is proper if and only if:*

(a) *the coefficients are integers,*

(b) *(xx) and (lx) are absolute invariants, and*

[2] The convention for order of multiplication is $L(C)$ followed by $L(I)$.

(c) *p_0' has the same sign as p_0 if $\{p\}$ is a solution of* (3) *and* $L\{p\}=\{p'\}$.

The necessity of conditions (a), (b) is well known [9]. Now a geometric $L(C)$ may be expressed as the product of linear substitutions which are permutations of $x_1, \cdots, x_\rho$, and of

$$A_{123}: \quad \begin{aligned} x_0' &= 2x_0 - x_1 - x_2 - x_3, \\ x_i' &= x_i + (x_0 - x_1 - x_2 - x_3), \qquad i = 1, 2, 3, \\ x_j' &= x_j, \qquad\qquad\qquad\qquad j = 4, \cdots, \rho. \end{aligned}$$

By Lemma 1 it is clear that under A_{123} any solution $\{p\}$ of (3) with $p_0 \geqq 0$ goes into a $\{p'\}$ of $p_0' \geqq 0$. Since A_{123} is an involution, it must then send a $\{p\}$ of $p_0<0$ into a $\{p'\}$ of $p_0'<0$. $L(C)$ must then have the same property.

The proof of the sufficiency of the conditions depends on the following theorem:

THEOREM 2 [7]. *Let $\{\gamma\}$ be an integer solution of equations* (1) *arranged so that $\gamma_1 \geqq \gamma_2 \geqq \cdots \geqq \gamma_\rho$. Further, let $\{\gamma\}$ satisfy the finite set of inequalities $p_0\gamma_0 - p_1\gamma_1 - \cdots - p_\rho\gamma_\rho \geqq 0$, where the characteristics $\{p\}$ run over the finite set of all proper solutions of* (3) *with $p_0<\gamma_0$ and so ordered that $p_1 \geqq p_2 \geqq \cdots \geqq p_\rho$. Then $\{\gamma\}$ is the characteristic of a homaloidal net.*

The ordering of $\{\gamma\}$ and $\{p\}$ is not necessary, but is stated for emphasis.

To demonstrate that an L of the form (2) satisfying (a), (b), (c) is proper, note first that $\{n; r_1, \cdots, r_\rho\}$ is an integer solution of (1) as a consequence of (a), (b), and (i). The invariance of the sign of p_0 assures that the inequalities of Theorem 2 hold. Thus $\{n; r_1, \cdots, r_\rho\}$ is the characteristic of a homaloidal net. By Lemma 2, the linear substitution L is proper.

The restriction of integer coefficients is indeed essential. An example has been exhibited [6, p. 863] for $\rho=9$ which satisfies all other conditions and is not an $L(C)$ since the numbers s_i, α_{ij} are rational. The proof of the sufficiency of the conditions of Theorem 1 could be proved by the method Coolidge uses; under the conditions given here, that argument is valid.

3. Some results.

COROLLARY 1. *If a solution $\{p\}$ of equations* (3) *satisfies*

$$p_0c_0 - p_1c_1 - \cdots - p_\rho c_\rho \geqq 0$$

for a single proper $\{c\}$ *which is a solution of* (1), *then it satisfies that relation for every proper* $\{c\}$.

COROLLARY 2. *Given solutions* $\{c\}$, $\{p\}$ *of* (1) *and* (3) *such that* $p_0>0$, $p_0c_0-p_1c_1-\cdots-p_nc_0<0$, *then* $\{c\}$ *is not proper.*

COROLLARY 3. *The linear substitution group generated by* A_{123} *and the permutations of* $x_1, x_2, \cdots, x_n$ *is completely characterized by conditions* (a), (b), (c) of Theorem 1.

The above results follow easily from Theorem 1. A similar theorem for characteristics for which x_0 never vanishes would be useful.

BIBLIOGRAPHY

1. A. B. Coble, Amer. Math. Monthly vol. 39 (1932) p. 293.
2. ———, *Cremona's diophantine equations*, Amer. J. Math. vol. 56 (1934) p. 459.
3. ———, *Algebraic geometry and theta functions*, Amer. Math. Soc. Colloquium Publications, vol. 10, New York, 1929, p. 11.
4. J. L. Coolidge, *Algebraic plane curves*, Oxford, 1931, p. 487.
5. G. B. Huff, *A note on Cremona transformations*, Proc. Nat. Acad. Sci. U.S.A. vol. 20 (1934) p. 428.
6. ———, *A geometry associated with Cremona's equations*, Amer. J. Math. vol. 62 (1940) pp. 865 and 863.
7. ———, *A sufficient condition that a C-characteristic be geometric*, Proc. Nat. Acad. Sci. U.S.A. vol. 29 (1943) p. 198.
8. S. Kantor, *Premiers fondements pour une theorie des transformations univoques*, Naples, 1891, p. 297.
9. National Research Council Report, *Selected topics in algebraic geometry*, Bulletin No. 63, Washington, D. C., 1929, p. 86.

SOUTHERN METHODIST UNIVERSITY

CONJUGATE NETS IN ASYMPTOTIC PARAMETERS

JANET MacDONALD

1. **Introduction.** Conjugate nets on an analytic nonruled surface in ordinary space may be studied either in conjugate parameters or in asymptotic parameters. The purpose of this paper is to make some contributions to the theory of conjugate nets in asymptotic parameters.

M. L. MacQueen has studied the pencil of quadrics each of which has second-order contact with a surface at a point and third-order contact with both curves of a conjugate net at the point. This pencil of quadrics is contained in the bundle of quadrics each of which has contact of at least the third order with both curves of a conjugate net at a point. The equation of this bundle of quadrics is deduced and the bundle is studied in this paper.

W. M. Davis in his Chicago doctoral dissertation, *Contributions to the theory of conjugate nets*, defined and studied several canonical configurations, considering the conjugate net as parametric. This paper presents some of Davis's results translated into asymptotic parameters.

A summary of that portion of the theory of conjugate nets in asymptotic parameters which is used in subsequent sections is given in §2. §3 deals with the bundle of quadrics mentioned above. Certain polar relations with respect to the quadrics of this bundle are presented in §4. In §5 certain loci and envelopes which arise in the study of a pencil of conjugate nets at a point on a surface are studied. §6 presents a study of Davis's canonical configurations in asymptotic parameters.

2. **Analytic basis.** The purpose of this section is to summarize[1] for later use portions of the theory of conjugate nets on an analytic nonruled surface referred to its asymptotic net in ordinary space.

Let the projective homogeneous coordinates $x_1, \cdots, x_4$ of a point P_x on a nonruled surface S in ordinary space be given by the parametric vector equation

$$x = x(u, v). \tag{2.1}$$

A necessary and sufficient condition that the asymptotic net be para-

Presented to the Society, November 26, 1943; received by the editors March 6, 1944.

[1] E. P. Lane, *A treatise on projective differential geometry*, Chicago, 1942, pp. 89, 105–113, 115–120, 125–129, 150–152, 180–190.

metric is that the point P_x satisfy two partial differential equations of the form

$$x_{uu} = px + \theta_u x_u + \beta x_v, \qquad x_{vv} = qx + \gamma x_u + \theta_v x_v, \qquad (\theta = \log \beta\gamma). \tag{2.2}$$

The coefficients are functions of u, v which are not arbitrary but must satisfy certain integrability conditions which need not be given here.

By means of equations (2.2) the third derivatives of x may be expressed as linear combinations of x, x_u, x_v, and x_{uv}.

A conjugate net N_λ on an integral surface of equations (2.2) can be represented by a curvilinear differential equation of the form

$$dv^2 - \lambda^2 du^2 = 0 \qquad (\lambda \neq 0), \tag{2.3}$$

where λ is a function of u, v. The two curves of the net N_λ through the point P_x will be denoted by C_λ and $C_{-\lambda}$ according as du/dv has value λ or $-\lambda$.

By the usual method, power-series expansions for the local coordinates of a point X sufficiently near a fixed point P_x and on the curve C_λ through the point P_x may be obtained:

$$\begin{aligned}
x_1 &= 1 + 2^{-1}(p + q\lambda^2)\Delta u^2 + 6^{-1}[(p_u + p\theta_u) + 3(p_v + \beta q)\lambda \\
&\quad + 3(q_u + \gamma p)\lambda^2 + (q_v + q\theta_v)\lambda^3 + 3q\lambda\lambda']\Delta u^3 + \cdots, \\
x_2 &= \Delta u + 2^{-1}(\theta_u + \gamma\lambda^2)\Delta u^2 + 6^{-1}[(p + \theta_u^2 + \theta_{uu}) + 3\kappa\lambda \\
&\quad + 3\chi\lambda^2 + (\gamma_v + \gamma\theta_v)\lambda^3 + 3\gamma\lambda\lambda']\Delta u^3 + \cdots, \\
x_3 &= \lambda\Delta u + 2^{-1}(\beta + \theta_v\lambda^2 + \lambda')\Delta u^2 + 6^{-1}[(\beta_u + \beta\theta_u) + 3\pi\lambda \\
&\quad + 3\kappa\lambda^2 + (q + \theta_v^2 + \theta_{vv})\lambda^3 + 3\theta_v\lambda\lambda' + \lambda'']\Delta u^3 + \cdots, \\
x_4 &= \lambda\Delta u^2 + 6^{-1}(\beta + 3\theta_u\lambda + 3\theta_v\lambda^2 + \gamma\lambda^3 + 3\lambda')\Delta u^3 + \cdots,
\end{aligned} \tag{2.4}$$

where

$$\lambda' = \lambda_u + \lambda\lambda_v, \qquad \kappa = \beta\gamma + \theta_{uv},$$
$$\pi = p + \beta_v + \beta\theta_v, \qquad \chi = q + \gamma_u + \gamma\theta_u.$$

The power series expansions for $C_{-\lambda}$ are obtained from these by first replacing λ' by $\lambda_u + \lambda\lambda_v$ and then changing the sign of λ.

The equation of the associate conjugate net is known to be

$$dv^2 + \lambda^2 du^2 = 0. \tag{2.5}$$

Power series expansions for a point X on a curve of the associate conjugate net are obtained from equations (2.4) by changing the sign of λ^2.

The equation of the pencil of conjugate nets determined by the net (2.3) is

$$dv^2 - \lambda^2 h^2 du^2 = 0, \tag{2.6}$$

where h is a constant independent of u and v.

Two types of congruences, Γ_1 and Γ_2, will be referred to in this paper. A line l_1 at a point P_x on a surface[2] may be regarded as determined by the points $x(1, 0, 0, 0)$ and $y(0, -a, -b, 1)$ and also by the planes

$$x_3 + bx_4 = 0, \qquad x_2 + ax_4 = 0. \tag{2.7}$$

If a and b are functions of u and v, the lines l_1 constitute a congruence Γ_1. The reciprocal line l_2 is the polar line of the line l_1 with respect to the quadric of Lie and may be regarded as determined by the points $\rho(-b, 1, 0, 0)$ and $\sigma(-a, 0, 1, 0)$ and also by the planes

$$x_4 = 0, \qquad x_1 + bx_2 + ax_3 = 0. \tag{2.8}$$

As in the case of a line l_1, if a and b are functions of u and v, the lines l_2 constitute a congruence Γ_2.

The axis at a point P_x of the conjugate net N_λ is a line l_1 for which a and b are denoted throughout this paper by a_a and b_a, respectively, and so are given by the formulas

$$a_a = 2^{-1}(\theta_v + \lambda_v/\lambda + \beta/\lambda^2), \quad b_a = 2^{-1}(\theta_u - \lambda_u/\lambda + \gamma\lambda^2). \tag{2.9}$$

The ray at a point P_x of the conjugate net N_λ is a line l_2 for which a and b are hereinafter denoted by a_r and b_r, respectively, and are given by the formulas

$$a_r = 2^{-1}(\theta_v + \lambda_v/\lambda - \beta/\lambda^2), \quad b_r = 2^{-1}(\theta_u - \lambda_u/\lambda - \gamma\lambda^2). \tag{2.10}$$

The principal lines of the two curves of the conjugate net N_λ at a point P_x are the associate conjugate tangents and have the equations

$$x_4 = 0, \qquad x_3^2 + \lambda^2 x_2^2 = 0. \tag{2.11}$$

The principal join is a line l_2 for which a and b are hereinafter denoted by a_j and b_j, respectively, and are given by the formulas

$$a_j = 2^{-1}(\theta_v + \lambda_v/\lambda + 5\beta/3\lambda^2), \quad b_j = 2^{-1}(\theta_u - \lambda_u/\lambda + 5\gamma\lambda^2/3). \tag{2.12}$$

The points

$$(-b_j - ia_j\lambda, 1, i\lambda, 0), \quad (-b_j + ia_j\lambda, 1, -i\lambda, 0) \tag{2.13}$$

are the principal points.

[2] Ibid., pp. 150–152.

The equations of the ray-point cubic at a point of the conjugate net N_λ are

$$lx_2x_3 - \beta x_2^3 - \gamma x_3^3 = 0, \qquad x_4 = 0, \tag{2.14}$$

where l is defined by

$$l = 2x_1 + (\theta_u - \lambda_u/\lambda)x_2 + (\theta_v + \lambda_v/\lambda)x_3. \tag{2.15}$$

The equations of the ray conic determined by the conjugate net N_λ at the point P_x are

$$4\beta\gamma x_2x_3 - l^2 = 0, \qquad x_4 = 0. \tag{2.16}$$

3. **Bundle of quadrics.** This section is devoted to deducing the equation of the bundle (linear two-parameter family) of quadrics each of which has contact of at least the third order with the curves C_λ and $C_{-\lambda}$ at the point P_x and to finding all the cones in the bundle.

If the general quadric, whose equation is

$$\begin{aligned} Ax_1^2 + Bx_2^2 + Cx_3^2 + Dx_4^2 + Ex_1x_2 + Fx_1x_3 + Gx_1x_4 \\ + Hx_2x_3 + Ix_2x_4 + Jx_3x_4 = 0, \end{aligned} \tag{3.1}$$

is to pass through the point $P_x(1, 0, 0, 0)$, the coefficient A must be zero. Contact of at least the third order with the curves C_λ and $C_{-\lambda}$ at the point P_x may be imposed by demanding that equation (3.1) be identically satisfied as far as the terms in Δu^3 by the power-series expansions for the curve C_λ given in equations (2.4) and also by the corresponding expansions for the curve $C_{-\lambda}$. After this has been done and certain notational changes which need not be explicitly stated here have been made, equation (3.1) reduces to

$$\begin{aligned} A(x_2x_3 - x_1x_4 - 3^{-1}\gamma\lambda^2 x_2x_4 - 3^{-1}(\beta/\lambda^2)x_3x_4) \\ + B(2b_\alpha x_3x_4 - 2\lambda^2 a_\alpha x_2x_4 - \lambda^2 x_2^2 + x_3^2) + Cx_4^2 = 0. \end{aligned} \tag{3.2}$$

Thus the following theorem is proved.

THEOREM 1. *The bundle of quadrics each of which has contact of at least the third order with the curves C_λ and $C_{-\lambda}$ at the point P_x is represented by equation* (3.2).

If $B=0$, equation (3.2) reduces to the equation of a pencil of quadrics having also second-order contact with the surface S at P_x. This pencil of quadrics has been studied by MacQueen.[3]

[3] M. L. MacQueen, *Pencils of quadrics associated with a conjugate net*, Journal of the Tennessee Academy of Science vol. 15 (1940) p. 421.

The cones in the bundle of quadrics represented by equation (3.2) are found by setting the discriminant of the left member of this equation equal to zero. The resulting equation can be reduced to

$$A^2(A^2 + 4B^2\lambda^2) = 0. \tag{3.3}$$

Three cases arise according as $A=0$, $B\neq 0$, or $A\neq 0$, $A^2+4B^2\lambda^2=0$, or $A=0$, $B=0$.

In the first case, when $A=0$, $B\neq 0$, equation (3.2) becomes

$$B(x_3^2 - \lambda^2 x_2^2 - 2\lambda^2 a_\alpha x_2 x_4 + 2b_\alpha x_3 x_4) + Cx_4^2 = 0. \tag{3.4}$$

This equation represents a pencil of noncomposite cones with vertex at (1, 0, 0, 0) in case, further,

$$B(\lambda^2 a_\alpha^2 - b_\alpha^2) + C \neq 0. \tag{3.5}$$

If the inequality (3.5) becomes an equality, equation (3.4) represents the osculating planes of the curves C_λ and $C_{-\lambda}$ at the point P_α.

In the second case when $A\neq 0$, $A^2+4B^2\lambda^2=0$, there result two pencils of cones, the equation of one of which is

$$\begin{aligned} &A[x_2x_3 - x_1x_4 - (3^{-1}\gamma\lambda^2 + i\lambda a_\alpha)x_2x_4 \\ &- (3^{-1}\beta/\lambda^2 - (i/\lambda)b_\alpha)x_3x_4 - 2^{-1}i\lambda x_2^2 + 2^{-1}(i/\lambda)x_3^2] + Cx_4^2 = 0. \end{aligned} \tag{3.6}$$

The equation of the other pencil of cones is obtained from equation (3.6) by replacing i by $-i$. The cones represented by equation (3.6) touch the tangent plane $x_4=0$ in one of the associate conjugate tangents. Every one of the cones of this pencil has its vertex at one of the principal points. The other pencil of cones has the same relation to the other associate conjugate tangent and the other principal point.

In the third case, when $A=0$, $B=0$, equation (3.2) reduces to $x_4^2=0$. Hence the tangent plane counted twice must be considered as a cone in the bundle.

The tangent plane $x_4=0$ intersects the most general quadric of the bundle (3.2) in two lines whose equations are

$$x_4 = 0, \qquad Bx_3^2 + Ax_2x_3 - B\lambda^2 x_2^2 = 0. \tag{3.7}$$

The two lines will be distinct in case $A^2+4B^2\lambda^2\neq 0$. The two lines of each pair separate the associate conjugate tangents harmonically. If $A=0$, $B\neq 0$, the lines in which the tangent plane intersects the quadric are the conjugate tangents. If $A\neq 0$, $B=0$, the lines are the asymptotic tangents.

4. **Polar relations.** The purpose of this section is to discuss certain

polar relations which exist among the configurations studied in this paper with respect to certain quadrics of the bundle (3.2).

The equations of the polar line of a line l_1 with respect to the most general quadric of the bundle (3.2) are $x_4=0$ and

$$A[x_1+(b+3^{-1}\gamma\lambda^2)x_2+(a+3^{-1}\beta/\lambda^2)x_3] + 2B[\lambda^2(a_a-a)x_2-(b_a-b)x_3]=0. \tag{4.1}$$

This equation represents a pencil of lines with its center at the point

$$[\lambda^2(a-a_a)(a+3^{-1}\beta/\lambda^2)+(b-b_a)(b+3^{-1}\gamma\lambda^2), \quad b_a-b,\ \lambda^2(a_a-a),\ 0]. \tag{4.2}$$

Thus we prove the following theorem.

THEOREM 2. *The polar lines of a line l_1 with respect to the quadrics of the bundle* (3.2) *form a pencil of lines in the tangent plane with the center at the point* (4.2).

For the axis, equation (4.1) reduces to

$$x_1+b_fx_2+a_fx_3=0. \tag{4.3}$$

Inspection of this equation leads to the following theorem.

THEOREM 3. *At a point P_x on the conjugate net N_λ the principal join is the polar line of the axis with respect to any quadric of the bundle* (3.2).

MacQueen proved[4] this theorem for any quadric of the pencil of quadrics studied by him.

Elimination of the parameters A and B from the equations of the polar line of a line l_2 with respect to the most general quadric of the bundle (3.2) yields the equation

$$\lambda^2x_2^2+x_3^2+[\lambda^2a_a(a-3^{-1}\beta/\lambda^2)+b_a(b-3^{-1}\gamma\lambda^2)]x_4^2 + 2^{-1}\lambda^2(\theta_v+\lambda_v/\lambda+3^{-1}\beta/\lambda^2+2a)x_2x_4 + 2^{-1}(\theta_u-\lambda_u/\lambda+3^{-1}\gamma\lambda^2+2b)x_3x_4=0. \tag{4.4}$$

Thus we prove the following theorem.

THEOREM 4. *The locus of the polar line of a line l_2 with respect to the quadrics of the bundle* (3.2) *is the quadric cone represented by equation* (4.4).

It is evident that this cone is intersected by the tangent plane in the associate conjugate tangents.

[4] Ibid., p. 422.

For the ray, equation (4.4) becomes

$$
\begin{aligned}
&\lambda^2 x_2^2 + x_3^2 + 4^{-1}[\lambda^2(\theta_v + \lambda_v/\lambda + \beta/\lambda^2)(\theta_v + \lambda_v/\lambda - 5\beta/3\lambda^2) \\
&\quad + (\theta_u - \lambda_u/\lambda + \gamma\lambda^2)(\theta_u - \lambda_u/\lambda - 5\gamma\lambda^2/3)]x_4^2 \\
&\quad + \lambda^2(\theta_v + \lambda_v/\lambda - 3^{-1}\beta/\lambda^2)x_2x_4 \\
&\quad + (\theta_u - \lambda_u/\lambda - 3^{-1}\gamma\lambda^2)x_3x_4 = 0.
\end{aligned} \tag{4.5}
$$

Davis's canonical plane is defined[5] to be the plane determined by the axis and the associate axis. The equation of this plane in asymptotic parameters is

$$
\gamma\lambda^2 x_2 - (\beta/\lambda^2)x_3 - 2^{-1}[(\beta/\lambda^2)(\theta_u - \lambda_u/\lambda) - \gamma\lambda^2(\theta_v + \lambda_v/\lambda)]x_4 = 0. \tag{4.6}
$$

This plane intersects the cone (4.5) in the axis and in a line l_1 which is the reciprocal polar of the associate principal join. Likewise the cone which is the locus of the polar line of the associate ray with respect to the quadrics of the bundle for the associate conjugate net intersects Davis's canonical plane in the associate axis and in the reciprocal polar of the principal join.

The polar planes of a point with respect to all the quadrics of a bundle are known[6] to pass through a point. The point through which pass the polar planes of a point P_y with respect to all the quadrics of the bundle (3.2) is found by setting equal to zero the coefficients of A, B, and C in the equation of the polar plane of P_y with respect to the most general quadric of the bundle (3.2) and solving for the ratios of $x_1, \cdots, x_4$. Two cases arise according as $y_4 \neq 0$ or $y_4 = 0$.

When $y_4 \neq 0$, the resulting coordinates are given by

$$
\begin{aligned}
x_1 &= \lambda^2 y_2^2 + y_3^2 - 3^{-1}(\gamma\lambda^2 b_u + \beta a_v)y_4^2 \\
&\quad + 2^{-1}\lambda^2(\theta_v + \lambda_v/\lambda + 3^{-1}\beta/\lambda^2)y_2y_4 \\
&\quad + 2^{-1}(\theta_u - \lambda_u/\lambda + 3^{-1}\gamma\lambda^2)y_3y_4,
\end{aligned} \tag{4.7}
$$

$$
x_2 = y_4(y_2 + b_u y_4), \quad x_3 = \lambda^2 y_4(y_2 + a_v y_4), \quad x_4 = 0.
$$

Hence, for a point P_y not in the tangent plane the polar planes with respect to all quadrics of the bundle (3.2) pass through a point in the tangent plane with coordinates (4.7). If the point P_y lies on the axis, the point (4.7) is indeterminate and all the polar planes pass through the principal join.

[5] W. M. Davis, *Contributions to the theory of conjugate nets*, doctoral dissertation, Chicago, 1932, p. 18.

[6] Virgil Snyder and C. H. Sisam, *Analytic geometry of space*, New York, 1914, p. 171.

When $y_4=0$, that is, when P_y (supposed distinct from the point P_x) lies in the tangent plane, the point through which pass the polar planes of P_y with respect to all the quadrics of the bundle (3.2) is indeterminate. The polar planes form a pencil whose axis may be represented by the equations

$$
\begin{aligned}
y_3x_2 + y_2x_3 + (y_1 + 3^{-1}\gamma\lambda^2 y_2 + 3^{-1}(\beta/\lambda^2)y_3)x_4 = 0,\\
\lambda^2 y_3x_2 - y_3x_3 + (\lambda^2 a_v y_2 - b_v y_3)x_4 = 0.
\end{aligned} \tag{4.8}
$$

If the point P_y lies on one of the associate conjugate tangents, the line (4.8) is the other associate conjugate tangent. If the point P_y coincides with P_x, the polar plane with respect to any quadric of the bundle is the tangent plane $x_4=0$.

5. Theory of a pencil of conjugate nets. This section deals with certain loci and envelopes which may be defined at a point of the surface in connection with a pencil of conjugate nets. The locus of the principal points, the envelope of the principal join, the envelope of the cone (4.5), the envelope of any particular one of the quadrics of the bundle (3.2), the envelope of Davis's canonical plane, the locus of Davis's canonical point, and the envelope of Davis's canonical quadric for nets of the pencil are studied.

The coordinates of one of the principal points for a general net of the pencil (2.6) are

$$
\begin{aligned}
x_1 = &- 2^{-1}(\theta_u - \lambda_u/\lambda + 5\gamma\lambda^2h^2/3)\\
&- 2^{-1}i\lambda h(\theta_v + \lambda_v/\lambda + 5\beta/3\lambda^2h^2),\\
x_2 = 1, \qquad x_3 = i\lambda h, \qquad x_4 = 0.
\end{aligned} \tag{5.1}
$$

Homogeneous elimination of h leads to the equations

$$
x_4 = 0, \qquad lx_2x_3 - (5/3)(\beta x_2^3 + \gamma x_3^3) = 0, \tag{5.2}
$$

where l is defined by equation (2.15). The same equation is obtained if the other principal point is used. Equations (5.2) represent a cubic, which will be called the *principal cubic*. This result may be stated in the following theorem.

THEOREM 5. *At a point on a surface the locus of the principal points of all the curves of a pencil* (2.6) *of conjugate nets that pass through the point is the principal cubic of the pencil, represented by equations* (5.2).

Comparison of equations (5.2) with equations (2.14) shows that the principal cubic and the ray-point cubic are in the pencil of cubics, in the tangent plane, with equations $x_4=0$ and

$$lx_2x_3 - k(\beta x_2^3 + \gamma x_3^3) = 0, \tag{5.3}$$

where k is an arbitrary parameter. The principal cubic is characterized by $k=5/3$, and the ray-point cubic by $k=1$.

The principal join for a general net of the pencil (2.6) is represented by the equations $x_4=0$ and

$$\begin{aligned} &x_1 + 2^{-1}(\theta_u - \lambda_u/\lambda + 5\gamma\lambda^2h^2/3)x_2 \\ &\quad + 2^{-1}(\theta_v + \lambda_v/\lambda + 5\beta/3\lambda^2h^2)x_3 = 0. \end{aligned} \tag{5.4}$$

Application of the theory of envelopes leads to the equations

$$x_4 = 0, \qquad 100\beta\gamma x_2x_3/9 - l^2 = 0. \tag{5.5}$$

This equation represents a conic, which will be called the *principal conic* of the pencil (2.6) of conjugate nets at the point P_x. This result may be stated in the following theorem.

THEOREM 6. *At a point on a surface the envelope of the principal joins of all the nets of a pencil* (2.6) *of conjugate nets is the principal conic, represented by equations* (5.5).

Comparison of equations (5.5) with equations (2.16) shows that the principal conic and the ray conic are in the pencil of conics, in the tangent plane, with equations

$$x_4 = 0, \qquad 4k\beta\gamma x_2x_3 - l^2 = 0, \tag{5.6}$$

where k is an arbitrary parameter. The principal conic is characterized by $k=25/9$, and the ray conic by $k=1$.

Replacing λ by λh in equation (4.5) and applying the theory of envelopes leads to the following equation of the envelope of the cone (4.5) for the pencil of conjugate nets (2.6):

$$\begin{aligned} &5\beta^2x_4^2m^3/3 + m^2n^2 + 5\gamma^2x_4^2n^3/3 + 25\beta^2\gamma^2x_4^4mn/8 \\ &\qquad - 625\beta^4\gamma^4x_4^8/768 = 0, \end{aligned} \tag{5.7}$$

where m and n are polynomials in x_2, x_3, and x_4 of degree four. Hence the envelope of the cone (4.5) is a surface of order eight.

Application of the theory of envelopes to the equation of any particular one of the quadrics of the bundle (3.2) leads to the following equation:

$$\begin{aligned} &(4A\beta x_3x_4/9)\{x_2x_4[A\gamma + 3B(\theta_v + \lambda_v/\lambda)] + 3Bx_2^2 - 3B\gamma x_3x_4\} \\ &\quad = \{A(x_2x_3 - x_1x_4) + B[x_3^2 - \beta x_2x_4 \\ &\qquad + (\theta_u - \lambda_u/\lambda)x_3x_4] + Cx_4^2\}^2. \end{aligned} \tag{5.8}$$

Hence, the envelope of any particular one of the quadrics of the bundle (3.2) is a quartic surface (5.8).

For a general net of the pencil (2.6) of conjugate nets the equation of Davis's canonical plane is

$$\gamma\lambda^2h^2x_2 - (\beta/\lambda^2h^2)x_3 - 2^{-1}[(\beta/\lambda^2h^2)(\theta_u - \lambda_u/\lambda) - \gamma\lambda^2h^2(\theta_v + \lambda_v/\lambda)]x_4 = 0. \tag{5.9}$$

Inspection shows that this plane passes through the cusp-axis,

$$2x_3 + (\theta_u - \lambda_u/\lambda)x_4 = 0, \quad 2x_2 + (\theta_v + \lambda_v/\lambda)x_4 = 0. \tag{5.10}$$

Thus we prove the following theorem.

THEOREM 7. *At a point on the surface Davis's canonical planes with respect to all the nets of the pencil* (2.6) *form a pencil of planes with the cusp-axis as axis.*

For a general net of the pencil (2.6) the coordinates of Davis's canonical point,[7] the intersection of the ray and the associate ray, are given by

$$\begin{aligned} x_1 &= (2^{-1}\beta/\lambda^2h^2)(\theta_u - \lambda_u/\lambda) - 2^{-1}\gamma\lambda^2h^2(\theta_v + \lambda_v/\lambda), \\ x_2 &= -\beta/\lambda^2h^2, \qquad x_3 = \gamma\lambda^2h^2, \qquad x_4 = 0. \end{aligned} \tag{5.11}$$

Homogeneous elimination of the parameter h yields the equations of the flex-ray:

$$x_4 = 0, \quad x_1 + 2^{-1}(\theta_u - \lambda_u/\lambda)x_2 + 2^{-1}(\theta_v + \lambda_v/\lambda)x_3 = 0. \tag{5.12}$$

Thus we prove the theorem.

THEOREM 8. *At a point on a surface the locus of Davis's canonical point with respect to a pencil* (2.6) *of conjugate nets on the surface is the flex-ray of the pencil of conjugate nets.*

Davis's canonical quadric at a point P_x of a conjugate net is defined[8] as the unique quadric surface having the following properties:

1. The quadric has second-order contact with the surface sustaining the net.

2. The axis and the ray of the net at P_x are reciprocal polars with respect to the quadric.

3. The quadric passes through the point P_y which is the harmonic conjugate of P_x with respect to the focal points of the axis.

[7] Davis, op. cit., p. 17.

[8] Ibid., p. 11.

The equation of Davis's canonical quadric at a point P_s of a general net of the pencil (2.6) is

$$
\begin{aligned}
&x_2x_3 - x_1x_4 + \gamma\lambda^2h^2x_2x_4 + (\beta/\lambda^2h^2)x_3x_4 \\
&\quad + [-2^{-1}\theta_{uv} + 4^{-1}\gamma_v\lambda^2h^2 + 4^{-1}\beta_u/\lambda^2h^2 \\
&\quad + 2^{-1}\gamma\lambda^2h^2(\theta_v + 2\lambda_v/\lambda) \\
&\quad + (2^{-1}\beta/\lambda^2h^2)(\theta_u - 2\lambda_u/\lambda)]x_4^2 = 0.
\end{aligned} \tag{5.13}
$$

Application of the theory of envelopes yields an equation of a quartic surface as the equation of the envelope of Davis's canonical quadrics for all the nets of a pencil (2.6) of conjugate nets.

6. **Canonical configurations.** In W. M. Davis's dissertation, *Contributions to the theory of conjugate nets*, the conjugate net is considered as parametric. The purpose of this section is to exhibit some of Davis's results in asymptotic parameters and to show some relationships that exist among the configurations studied by him. The canonical configurations referred to are those defined[9] by Davis.

The equation of Davis's canonical quadric for the net N_λ may be obtained from equation (5.13) by setting $h=1$. The following equation of his associate canonical quadric may be obtained from the same equation (5.13) by setting $h=1$ and replacing λ^2 by $-\lambda^2$:

$$
\begin{aligned}
&x_2x_3 - x_1x_4 - \gamma\lambda^2x_2x_4 - (\beta/\lambda^2)x_3x_4 \\
&\quad + [-2^{-1}\theta_{uv} - 4^{-1}\gamma_v\lambda^2 - 4^{-1}\beta_u/\lambda^2 - 2^{-1}\gamma\lambda^2(\theta_v + 2\lambda_v/\lambda) \\
&\quad - (2^{-1}\beta/\lambda^2)(\theta_u - 2\lambda_u/\lambda)]x_4^2 = 0.
\end{aligned} \tag{6.1}
$$

If the equations of Davis's canonical quadric and associate canonical quadric are added, and if the result is divided by 2, the equation of the quadric of Wilczynski is obtained:

$$x_2x_3 - x_1x_4 - 2^{-1}\theta_{uv}x_4^2 = 0. \tag{6.2}$$

Hence the quadric of Wilczynski belongs to the pencil of quadrics determined by Davis's canonical quadric and associate canonical quadric.

The coordinates of Davis's canonical point for the net N_λ are given by

$$
\begin{aligned}
&x_1 = (2^{-1}\beta/\lambda^2)(\theta_u - \lambda_u/\lambda) - 2^{-1}\gamma\lambda^2(\theta_v + \lambda_v/\lambda), \\
&x_2 = -\beta/\lambda^2, \qquad x_3 = \gamma\lambda^2, \qquad x_4 = 0.
\end{aligned} \tag{6.3}
$$

A little computation shows that the canonical point (6.3) and the

[9] Ibid., pp. 17–25.

canonical plane (4.6) are pole and polar with respect to both the canonical quadric and the associate canonical quadric.

The pencil of lines in the canonical plane with center at P_x is defined by Davis to be the first canonical pencil of the net, and the pencil of lines in the tangent plane at P_x with center at the canonical point is defined to be the second canonical pencil of the net. The flex-ray, represented by equations (5.12), belongs to the second canonical pencil.

The intersection of the tangent plane with the canonical plane is defined by Davis to be the first canonical tangent. The equations of the first canonical tangent in asymptotic parameters are $x_4=0$ and

$$\gamma\lambda^2 x_2 - (\beta/\lambda^2)x_3 = 0. \tag{6.4}$$

The line joining the canonical point to the point P_x is defined by Davis to be the second canonical tangent. The equations of this line are $x_4=0$ and

$$\gamma\lambda^2 x_2 + (\beta/\lambda^2)x_3 = 0. \tag{6.5}$$

Davis's thesis includes a study of congruences of pairs of lines which are polar lines with respect to his canonical quadric at a point of a conjugate net of a surface, when one of the lines, l_1, passes through P_x but does not lie in the tangent plane, and the other line, l_2, lies in the tangent plane but does not pass through P_x. Only such reciprocal polar lines as belong to Davis's canonical pencils are considered.

As in §2, a line l_1 may be represented by equations

$$x_2 + ax_4 = 0, \qquad x_3 + bx_4 = 0. \tag{6.6}$$

The line l_2 which is the reciprocal polar of l_1 with respect to Davis's canonical quadric may be represented by equations $x_4=0$ and

$$x_1 + (b - \gamma\lambda^2)x_2 + (a - \beta/\lambda^2)x_3 = 0. \tag{6.7}$$

The lines l_1 and l_2 will be lines of the first and the second canonical pencils (in the sense of Davis), respectively, if

$$a = 2^{-1}(\theta_v + \lambda_v/\lambda + h\beta/\lambda^2), \quad b = 2^{-1}(\theta_u - \lambda_u/\lambda + h\gamma\lambda^2), \tag{6.8}$$

where h is a parameter.

If $h=1$, then $a=a_a$, $b=b_a$, and $a-\beta/\lambda^2=a_r$, $b-\gamma\lambda^2=b_r$. Hence, for $h=1$, the line l_1 is the axis and l_2 the ray of the conjugate net N_λ. If $h=-1$, then l_1 is the associate axis. If $h=3$, then l_2 is the associate ray. If $h=2$, then l_2 is the flex-ray. If $h=0$, then l_1 is the cusp-axis. If $h=\infty$, the lines l_1 and l_2 are Davis's first and second canonical tangents, respectively. Any pair of lines l_1 and l_2 given by the same

value of h in (6.8) are reciprocal polars with respect to Davis's canonical quadric. It is seen that the associate axis and the associate ray are not corresponding lines. Likewise, the flex-ray and the cusp-axis are not corresponding lines.

Davis defines at each point of a surface three tangents analogous to the tangents of Darboux. All the quadric surfaces represented by the equation

$$x_2x_3 - x_1x_4 + \gamma\lambda^2 x_2x_4 + (\beta/\lambda^2)x_3x_4 + k_4x_4^2 = 0 \tag{6.9}$$

have second order contact with the surface at P_x and intersect the surface in a curve with a triple point at P_x. The triple point tangents of this curve are defined to be *the triple tangents* of the conjugate net at the point P_x. They can be shown to intersect the ray-point cubic (2.14) in three collinear points lying on a line in Davis's second canonical pencil, defined by Davis to be *the line of collineation* of the net N_λ. The associate line of collineation is the reciprocal polar of the associate axis with respect to Davis's canonical quadric.

In addition to the lines already mentioned, the principal join and the associate principal join are also members of Davis's second canonical pencil, the principal join being characterized by $h = 11/3$, and the associate principal join by $h = 1/3$.

UNIVERSITY OF CHICAGO

CONTRACTIONS IN NON-EUCLIDEAN SPACES

F. A. VALENTINE

The existence of an extension of the range of definition of a function $f(x)$ defined on a set S of a metric space M to a metric space M' so as to preserve a contraction of the type

$$\|f(x_1), f(x_2)\|' \leqq \|x_1, x_2\| \tag{1}$$

depends upon M and M'. The author has previously shown [3, 4][1] that for $M = M'$ the extension exists when M is: (1) the n-dimensional Euclidean space; (2) the surface of the n-dimensional Euclidean sphere; (3) the general Hilbert space. In this brief article *the extension is shown to exist when each M and M' is the n-dimensional hyperbolic space.* The method used to prove this result is applied to a metric space which includes both the hemispherical and hyperbolic cases. Hence a unification of results is also obtained.

As shown in the previous papers [3, 4] a necessary and sufficient condition for a contraction to be extensible in M and M' is the property E, which is restated as follows.

PROPERTY E. *Consider in each of the metric spaces M and M' a set of spheres, such that to each sphere $S_i \in M$, having center x_i and radius r_i, there corresponds a sphere $S_i' \in M'$, having center x_i' and radius r_i'. Furthermore suppose that*

$$\begin{gathered} r_i = r_i', \\ \|x_i', x_j'\|' \leqq \|x_i, x_j\| \end{gathered} \tag{2}$$

for all corresponding spheres S_i and S_i', and for all corresponding pairs (S_i, S_j) and (S_i', S_j').

The spaces M and M' are said to have the extensibility property E *if conditions* (2) *and*

$$\prod_i S_i \neq 0 \tag{3}$$

imply that

$$\prod_i S_i' \neq 0. \tag{4}$$

If the above statement holds for $M = M'$, the space M is said to have property E.

Presented to the Society, April 29, 1944; received by the editors February 18, 1944.

[1] Numbers in brackets refer to references at the end of the paper.

For convenience of discussion let M be an n-dimensional metric space which can be imbedded in an $(n+1)$-dimensional Euclidean space $\mathfrak{R}$. Let R_i be the Euclidean vector emanating from the origin of $\mathfrak{R}$ to the point $x_i \in M$. It is assumed that there exists a symmetric real-valued bilinear product $R_i \cdot R_j$ such that $R_i \cdot R_i = k^2 = \text{const.}$ defines the metric space M. Suppose that for any two points x_i and x_j in M, R_i and R_j determine a Euclidean plane which intersects M in a *unique* continuous curve joining x_i and x_j. This curve is defined to be a geodesic. Furthermore suppose the distance $\|x_i, x_j\|$ in M is defined to be

$$\|x_i, x_j\| = F(R_i \cdot R_j) \geqq 0, \tag{5}$$

where $F(u)$ is either a single-valued *increasing* function of u or a single-valued *decreasing* function of u.

THEOREM 1. *If the n-dimensional metric space M has the above properties, it possesses the property* E.

PROOF. To prove this we consider the case $F(u)$ is an *increasing* function of u. When $F(u)$ is decreasing the proof is obtained by a uniform change in the direction of the inequality signs. On account of a theorem of Helly[2] type [2, 1], to prove Theorem 1 it is sufficient to establish property E for $i = 1, \cdots, n+1$.

Let $\Delta(x_1, \cdots, x_{n+1})$ be the simplex (degenerate or nondegenerate) in M determined by the points x_i $(i = 1, \cdots, n+1)$. Condition (4) implies that $\Delta(x_{i_1}, x_{i_2}) \cdot S_{i_1} \cdot S_{i_2} \neq 0$. If we have $\Delta(x_{i_1}, \cdots, x_{i_{r+1}}) \cdot \prod_{j=i_1}^{i_r} S_j \neq 0$ $(i = 1, \cdots, n+1;\ 2 \leqq r \leqq n)$, then since by (4) $\prod_{j=i_1}^{i_{r+1}} S_j \neq 0$, the theorem of Helly[3] type implies in the r-dimensional subspace that $\Delta(x_{i_1}, \cdots, x_{i_{r+1}}) \cdot \prod_{j=i_1}^{i_{r+1}} S_j \neq 0$. Hence

$$\Delta(x_1, \cdots, x_{n+1}) \cdot \prod_{i=1}^{n+1} S_i \neq 0 \tag{6}$$

is established by induction. *Suppose* that $\Delta(x_1', \cdots, x_{n+1}')$ is *not covered* by the spheres S_i'. Then choose x and x' so that

$$x \in \Delta(x_1, \cdots, x_{n+1}) \cdot \prod_{i=1}^{n+1} S_i, \quad x' \in \Delta(x_1', \cdots, x_{n+1}') - \sum_{i=1}^{n+1} S_i', \tag{7}$$

and let R and R' be the Euclidean vectors emanating from the origin

[2] The theorem states: *If each $n+1$ sets of a family of closed bounded, convex sets of the n-dimensional Euclidean space intersect, then there is a point common to all the sets.* For a more general topological theorem of the same type, see Alexandroff and Hopf [1, p. 297].

[3] Loc. cit.

of $\mathfrak{R}$ to the points x and x' respectively. Conditions (2) and (5) yield the results

$$R_i \cdot R_j \geqq R_i' \cdot R_j' \qquad (i, j = 1, \cdots, n+1). \tag{8}$$

Also conditions (7) and the first of conditions (2) imply that $\|x_i', x'\| > \|x_i, x\|$. Hence by (5) we have

$$R' \cdot R_i' > R \cdot R_i. \tag{9}$$

Since the line of shortest length joining x_i and x_j lies in the plane determined by R_i and R_j, the simplex $\Delta(x_1, \cdots, x_{n+1})$ is contained in the smaller solid angle α determined by $R_1, \cdots, R_{n+1}$. Hence condition (7) implies that R lies inside the solid angle α. A corresponding statement with primes holds for R'. *Hence there exist* real constants a_i and a_i' such that

$$a_i \geqq 0, \qquad a_i' \geqq 0, \qquad \sum_{i=1}^{n+1} a_i \neq 0, \qquad \sum_{i=1}^{n+1} a_i' \neq 0,$$

and such that

$$R = a_i R_i, \qquad R' = a_i' R_i' \qquad (i \text{ summed}). \tag{10}$$

Multiplying (8) by $a_i a_j'$, summing on i and j, one obtains

$$(a_i R_i) \cdot (a_j' R_j) \geqq (a_i R_i') \cdot (a_j' R_j'),$$

whence by (10)

$$R \cdot (a_j' R_j) \geqq (a_i R_i') \cdot R'. \tag{11}$$

Similarly multiplying (9) by a_i, summing on i, we get

$$R' \cdot (a_i R_i') > R \cdot R. \tag{12}$$

Conditions (11) and (12) imply that

$$R \cdot (a_j' R_j) > R \cdot R. \tag{13}$$

However multiplying (9) by a_i', we get

$$R' \cdot R' = R' \cdot (a_i' R_i') > R \cdot (a_i' R_i). \tag{14}$$

Since $R \cdot R = R' \cdot R' = k^2$, conditions (13) and (14) are contradictory. Hence the assumption that $\Delta(x_1', \cdots, x_{n+1}')$ is not covered by the spheres S_i' is *false*. Since $\Delta(x_1', \cdots, x_{n+1}') \cdot S_i' \cdot S_j' \neq 0$, and since Δ is covered by the spheres S_i', a theorem[4] of Knaster, Kuratowski and

[4] See Alexandroff and Hopf [1, p. 377]. The theorem states: *If the closed sets A_i cover the simplex T, and if each side $a_{i_1} \cdots a_{i_r}$ of T is such that $a_{i_1} \cdots a_{i_r} \subset A_1 + \cdots + A_{i_r}$, then $A_1 \cdot A_2 \cdot \cdots \cdot A_{n+1} \neq 0$.*

Mazurkiewicz implies by induction that $\prod_{i=1}^{n+1} S_i' \neq 0$. Since condition (4) now holds for each set of $n+1$ of the spheres S_i', the theorem of Helly[5] type implies that (4) holds for all the spheres S_i'.

We now readily prove the following corollary.

COROLLARY 1. *The property* E *holds for the n-dimensional hyperbolic space.*

For the hyperbolic space M this corollary is an immediate consequence of the fact that M can be defined as the points $(x_1, x_2, \cdots, x_{n+1})$ in the $(n+1)$-dimensional Euclidean space which are on one sheet of the hyperboloid[6]

$$k^2 x_1^2 - x_2^2 - x_3^2 - \cdots - x_{n+1}^2 = k^2.$$

Here $R_i \cdot R_j$ is defined to be the bilinear form

$$R_i \cdot R_j \equiv k^2 x_{i1} x_{j1} - x_{i2} x_{j2} - \cdots - x_{in+1} x_{jn+1},$$

and $F(u) = k \cosh^{-1}(u/k^2)$. These have the properties required for the proof of Theorem 1. A similar argument holds for the open hemispherical case.

The extensibility of $f(x)$ to the whole space M so as to preserve condition (1) now follows as developed in the previous work of the author [3, pp. 105–106].

BIBLIOGRAPHY

1. Alexandroff and Hopf, *Topologie*, vol. 1, 1935, Julius Springer, Berlin.

2. E. Helly, Jber. Deutschen Math. Verein. vol. 32 (1923) pp. 175–176.

3. F. A. Valentine, *On the extension of a vector function so as to preserve a Lipschitz condition*, Bull. Amer. Math. Soc. vol. 49 (1943) pp. 100–108.

4. ———, *A Lipschitz condition preserving extension for a vector function*. To appear in Amer. J. Math.

5. Coxeter H., *Non-Euclidean geometry*, Toronto, 1942.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES

[5] Loc. cit.

[6] Coxeter [5, pp. 209, 248].

A GENERALIZATION OF MOORE'S THEOREM ON SIMPLE TRIODS

GAIL S. YOUNG, JR.

R. L. Moore has proved that the plane does not contain uncountably many mutually exclusive simple triods.[1] The generalization to spaces of higher dimension given below appears to have escaped notice, and to be of some interest.

DEFINITION. *If n is a non-negative integer, by a T_n-set we shall mean a continuum which is the sum of an n-cell, g, and an arc, t, such that $g \cdot t$ is a point which is an end point of t and a relatively interior point of g. The point $g \cdot t$ will be called a junction point.*

Obviously, a T_1-set is a simple triod.

THEOREM. *Euclidean n-space does not contain uncountably many mutually exclusive T_{n-1}-sets.*

PROOF. Supose that the theorem is false. Then for some positive number ϵ there exists an uncountable collection, G, of mutually exclusive T_{n-1}-sets such that the junction point of each is at distance greater than ϵ from the boundary of its $(n-1)$-cell. There is a point, P, which is a point of condensation of the set of all junction points of elements of G; let U be a spherical domain of n-space with center P and radius less than $\epsilon/2$. If the point X of U is a junction point of an element T of G, let $g(X)$ denote the component that contains X of the intersection of U and the $(n-1)$-cell of T. It is an easy consequence of the Alexander duality theorem that $g(X)$ separates U. Hence the collection, G', of all sets $g(X)$ is an uncountable collection of cuttings of U, and it is clearly non-separated. By a theorem due to Whyburn,[2] G' contains an uncountable saturated subcollection, G''. But if $g(X)$ is an element of G'', U contains an arc, t, which is in the T_{n-1}-set of G that contains X and which has only X in common with $g(X)$. Since the elements of G are mutually exclusive, no element of G'' separates a point of t from $g(X)$ in U, which is a contradiction.

PURDUE UNIVERSITY

Presented to the Society, April 29, 1944; received by the editors March 25, 1944.

[1] In fact, Moore has proved that the plane does not contain uncountably many mutually exclusive triodic continua. See *Concerning triods in the plane and the junction points of plane continua*, Proc. Nat. Acad. Sci. U.S.A. vol. 14 (1928) pp. 85–88, and *Concerning triodic continua in the plane*, Fund. Math. vol. 13 (1929) pp. 261–263.

[2] Theorem 2.2 of chap. 3 of his book, *Analytic topology*, Amer. Math. Soc. Colloquium Publications, vol. 28.

ON ROTATION GROUPS OF PLANE CONTINUOUS CURVES UNDER POINTWISE PERIODIC HOMEOMORPHISMS

DICK WICK HALL

In this paper we make use of the work of G. T. Whyburn[1] on light interior transformations and on orbit decompositions of certain spaces to obtain a theorem by means of which a certain subset of the orbits of points under a periodic transformation $T(M)=M$ may be given a linear ordering. This theorem is then used to obtain an accessibility theorem for plane continuous curves similar to one previously published by L. Whyburn.[2] We take this opportunity to express our indebtedness to G. E. Schweigert for suggesting the proof of Theorem I given below and thus eliminating the longer and less interesting proof previously obtained by the author. For any $x\in M$, the orbit of x under T means $O(x)=\sum_{i=-\infty}^{\infty}T^i(x)$.

THEOREM I. *Let M be a locally connected continuum (that is, a continuous curve) and $T(M)=M$ an arbitrary periodic homeomorphism. Then if a and b are arbitrary points of M lying in different orbits under T and if axb is any simple arc in M joining a and b, then there must exist a simple arc $a'x'b'$ in M lying in the orbit of axb under T such that a' belongs to $O(a)$, b' belongs to $O(b)$ and no two points of $a'x'b'$ lie in the same orbit under T. Furthermore, the point a' may be any arbitrary preassigned point of the orbit of a.*

Proof (Schweigert). Let M' be the hyperspace obtained by decomposing the space M into its orbits under T. Then, since the orbit decomposition is continuous,[3] it follows[4] that there exists a light interior transformation $f(M)=M'$, namely, the transformation given by and associated with the orbit decomposition. Let axb be the given arc in M. Then we may assume without loss of generality that axb has precisely the point a in common with $O(a)$ and precisely the point b in common with $O(b)$. Define $K=f(axb)$. Then K is a locally connected continuum containing $c=f(a)$ and $d=f(b)$. Let cyd be an arc in K joining c to d. Now let a' be an arbitrary point of $O(a)$. Then[5]

Presented to the Society, February 26, 1944; received by the editors April 13, 1944.

[1] See G. T. Whyburn, *Analytic topology*, Amer. Math. Soc. Colloquium Publications, vol. 28, 1942, pp. 182–189 and 239–262.

[2] See L. Whyburn, *Rotation groups about a set of fixed points*, Fund. Math. vol. 28 (1937) pp. 124–130, in particular p. 127.

[3] See G. T. Whyburn, loc. cit. p. 258.

[4] See G. T. Whyburn, loc. cit. p. 130.

[5] See G. T. Whyburn, loc. cit. p. 186.

there must exist a simple arc $a'x'b'$ in M such that $f(a'x'b')=cyd$ is topological. By definition of f we see that each point of $a'x'b'$ belongs to the orbit of some point of axb, and from the one-to-oneness of this transformation it is immediate that no two points of $a'x'b'$ lie in the same orbit under this transformation. This completes the proof.

COROLLARY. *The same conclusion holds for any pointwise periodic $T(M)=M$ if we impose the additional restriction either that T have equicontinuous powers*[6] *or that the period function remain bounded on the arc axb.*

The accessibility theorem for plane continuous curves mentioned in the introductory paragraph of this paper may be stated as follows.

THEOREM A (L. WHYBURN). *If M is a plane continuous curve and $T(M)=M$ is a homeomorphism and if C is an element of a rotation group of M under T of order at least two, then C has property S.*[7]

The object of our second theorem is to obtain a result similar to Theorem A, but with the emphasis in the hypothesis placed upon the type of the transformation T rather than upon the order of the rotation group under consideration. Before stating the theorem we recall certain important subsets of M. By L we denote the closed invariant subset of M consisting of those points at which the period function has an unbounded limit superior. By K we denote the collection of all fixed points of M under T. If R is a component of $M-K$, then R is an element of a rotation group under T; this rotation group consists exactly of the orbit of R under T; and its order is the number of components which it contains. The order of a rotation group under T may, of course, be either finite or infinite.

We are now in a position to state our second theorem.

THEOREM II. *Let M be a plane continuous curve and $T(M)=M$ an arbitrary homeomorphism, while R denotes an element of some rotation group of M under T. Then*

(a) *If K is locally connected then R has property S and every point of $F(R)=\overline{R}-R$ is regularly accessible*[8] *from R.*

(b) *If T is pointwise periodic and has equicontinuous powers we get the same conclusion as in* (a).

(c) *If T is pointwise periodic then every point of $F(R)$ which is not a point of L is regularly accessible from R.*

Proof. Note that (a) is immediate from a theorem of G. T. Why-

[6] See G. T. Whuburn, loc. cit. p. 258.
[7] See G. T. Whyburn, loc. cit. p. 20.
[8] See G. T. Whyburn, loc. cit. p. 111.

burn,[9] and that it also follows at once from the proof of Theorem A.

We give the proofs of (b) and (c) simultaneously, making use of the corollary to Theorem I in each case. It is to be noted that our proof is similar to the original proof of Theorem A.

Denote by d a positive number exceeding the diameter of the set M, and suppose that M is embedded in the upper half of the Euclidean plane. If the theorem be false there must exist a point p in $F(R)$ which is not regularly accessible from R and (unless T has equicontinuous powers) such that p does not lie in L. This means that there exists a positive number ϵ and a sequence of points $\{p_i\}$ of R converging to p such that no two points of this sequence may be joined in R by a connected set of diameter less than 17ϵ. Let C_p be a circle of radius 8ϵ having its center at the point p. No generality is lost by the following assertion:

(1) *For every i the set $O(p_i)$ lies within C_p; no two of the points p_i may be joined by a connected subset of R lying within this circle; and if T does not have equicontinuous powers then there exists an integer N such that no point of M lying within C_p has period greater than N under T.*

For some point q' of R exterior to C_p we construct arcs p_iq' in R for every integer i and we denote the first intersection of the arc p_iq' with the circle C_p by q_i. It follows from (1) that the arcs p_iq_i are pairwise disjoint and we may assume, exactly as in the proof of Theorem A, that this sequence of arcs converges to a limiting set H which is a subcontinuum of K. Making use of the corollary to Theorem I we can insure that no arc p_iq_i meets the orbit of any point of M in more than a single point, and that no two of these arcs meet the orbit of the same point of M. This means, in particular, that no two consecutive images under T of any one of these arcs will have a point in common. We may assume that the sequence $\{q_i\}$ converges monotonically on C_p to a point q.

We place the x-axis in such a position that p lies at the point $(-4\epsilon, 2d)$ and q at the point $(4\epsilon, 2d)$. By L_i $(i=\pm1, \pm2, \pm3)$ we denote the line segment joining $(i\epsilon, 0)$ to $(i\epsilon, 4d)$, and by $D_{i,j}$ the interior of the rectangle formed by L_i, L_j, $y=0$, and $y=4d$.

Using the fact that H is a subset of K and either (b) or (c)[10] we may make the following assumption without loss of generality:

(2) *If x is any point of any arc p_iq_i then $O(x)$ lies within a circle having its center at some point of H and diameter $\epsilon/2$.*

Let r_is_i be a subarc of p_iq_i with its interior in $D_{-3,3}$ and its end

[9] See G. T. Whyburn, *Concerning the open subsets of a plane continuous curve*, Proc. Nat. Acad. Sci. U.S.A. vol. 13 (1927) pp. 650–657, in particular Theorems 1 and 5 of this paper.

[10] See G. T. Whyburn, *Analytic topology*, loc. cit. p. 252.

points on L_{-3} and L_3; $r_i's_i'$ the arc $T(r_is_i)$; $x_i'y_i'$ a subarc of $r_i's_i'$ with its interior in $D_{-2,2}$ and its end points on L_{-2} and L_2; and x_iy_i the arc $T^{-1}(x_i'y_i')$. The existence of the arc $x_i'y_i'$ follows from (2) and by the same token we see that x_iy_i is a subarc of r_is_i having its end points in the respective regions $D_{-3,-1}$, $D_{1,3}$. Now the sequence of arcs $\{x_i'y_i'\}$ may be assumed to converge to a subcontinuum H' of H which is, of course, disjoint with every one of these arcs. This means that by taking a subsequence the following assumption will hold.

(3) *For any fixed integer i and every k exceeding i the arc $x_k'y_k'$ separates $D_{-2,2}$ between $x_i'y_i'$ and $x_n'y_n'$ for every n greater than k. Thus $x_k'y_k'$ separates $D_{-2,2}$ between $x_i'y_i'$ and H' for every k exceeding i.*

For any fixed value of i we know that the closed sets $O(x_iy_i)$ and H' are disjoint, which means that there will exist a region U_i in the plane containing H' but disjoint with $O(x_iy_i)$. We may assume that U_i contains $O(x_ky_k)$ for every k exceeding i. We also know, in view of this last remark, that for i fixed either x_iy_i separates $D_{-1,1}$ between $x_i'y_i'$ and every $x_k'y_k'$ for k exceeding i or $x_i'y_i'$ separates this region between x_iy_i and $x_k'y_k'$ for every k exceeding i. By taking a subsequence and renumbering we can insure that the same one of these two statements holds for every value of i and thus obtain the following assertion.

(4) *If i and k be any two distinct integers then in the region $D_{-1,1}$ the four arcs x_iy_i, $x_i'y_i'$, x_ky_k, $x_k'y_k'$ must occur either in the order just specified or in the alternative order $x_i'y_i'$, x_iy_i, $x_k'y_k'$, x_ky_k.*

No generality is lost by the assumption that for every integer i the arc x_iy_i has an interior point within the circle C_z, having its center at a point z on the y-axis and radius sufficiently small so that any two points of M lying within C_z may be joined by an arc of M the orbit of which lies within $D_{-1,1}$. This enables us to find a simple arc a_ib_k lying in M with $O(a_ib_k)$ in $D_{-1,1}$ and having exactly the points a_i, b_k in common with x_iy_i and x_ky_k, respectively. As the two arrangements given in (4) are symmetrical we need treat only the case of the first one; the other will follow by a simple interchange of the letters i and k. From (1) we see that the arc a_ib_k must contain at least one point of the closed set K, and we denote by f_{ik} the last point of K on this arc. Then if g_k be the first point of $O(x_ky_k)$ on the arc $f_{ik}b_k$, it follows that for some integer n the point $g_k' = T^n(g_k)$ lies on the arc $x_k'y_k'$. Thus the arc $f_{ik}g_k' = T^n(f_{ik}g_k)$ is a simple arc lying in $D_{-1,1}$ and joining the point f_{ik} to a point of $x_k'y_k'$, while containing no point of x_ky_k. This contradiction completes the proof of Theorem II.

UNIVERSITY OF MARYLAND

EQUIVALENCE OF A PROBLEM IN MEASURE THEORY TO A PROBLEM IN THE THEORY OF VECTOR LATTICES

GEORGE W. MACKEY

In connection with some other work of the author a question arises concerning the form of the general bounded linear[1] functional on certain vector lattices.[2] It is the purpose of this note to show that this question is completely equivalent to a question in measure theory which has been discussed and partially answered by Ulam [2].

Let S be an abstract set and let $\mathfrak{F}$ be the vector lattice of all real-valued functions defined on S. For each s_0 in S the function F on $\mathfrak{F}$ such that $F(f)=f(s_0)$ for all f in $\mathfrak{F}$ is clearly a linear functional. We shall call it the point functional belonging to s_0 or simply a point functional. Obviously every point functional and hence every finite linear combination of point functionals is bounded in the sense that it carries every bounded subset of $\mathfrak{F}$ into a bounded set of real numbers. Our question is as to whether every bounded linear functional on $\mathfrak{F}$ is a finite linear combination of point functionals. We shall show that this is the case if and only if there exists no countably additive measure α which is defined for all subsets of S, which is zero at points, which takes on only the values zero and one, and which does not vanish identically.

It is well known that every bounded linear functional on a vector lattice is a difference of non-negative linear functionals[3] and it is obvious that a non-negative linear functional is bounded. It follows that we need only consider non-negative linear functionals. Passages from a measure to a non-negative linear functional defined on a class of functions and vice versa are of frequent occurrence in mathematical literature. The proof of our theorem rests basically on the fact that when the methods used in effecting these passages are applied to the case at hand one obtains a natural one-to-one correspondence between the non-negative linear functionals on $\mathfrak{F}$ and the countably ad-

Received by the editors March 7, 1944.

[1] By a linear functional we mean a functional which is additive and homogeneous; that is, one which preserves linear combinations.

[2] See chap. 7 of [1] for definitions of the terms from the theory of vector lattices which we shall use. Numbers in brackets refer to the references cited at the end of the paper.

[3] This is proved on p. 115 of [1] for additive functionals and it is clear that an additive non-negative functional must be linear.

ditive measures which are defined on all subsets of S and take on only a finite number of different values.

LEMMA 1. *If P is a non-negative linear functional on $\mathfrak{F}$ and if $\phi_1, \phi_2, \cdots$ are the characteristic functions of mutually disjoint subsets of S then at most a finite number of the numbers $P(\phi_1), P(\phi_2), \cdots$ are different from zero.*

PROOF. Suppose the contrary and change the notation so that $P(\phi_1), P(\phi_2), \cdots$ are all different from zero. Let E_i be the set whose characteristic function is ϕ_i, $i=1, 2, \cdots$, and let $E=E_1 \cup E_2 \cup \cdots$. For each s in E_i let $g(s)=1/P(\phi_i)$ and for each s in $S-E$ let $g(s)=0$. Then for each $n=1, 2, \cdots$, $g \geqq (\phi_1/P(\phi_1))+(\phi_2/P(\phi_2))+\cdots+(\phi_n/P(\phi_n))$. Hence $P(g) \geqq 1+1+\cdots+1=n$ for $n=1, 2, \cdots$ and this is impossible.

LEMMA 2. *If P is a non-negative linear functional on $\mathfrak{F}$ then $P=P_1+P_2$ where P_1 is a finite linear combination of point functionals and P_2 is a non-negative linear functional which vanishes on characteristic functions of points.*

PROOF. By Lemma 1 there are at most a finite number of points in S whose characteristic functions are not taken into zero by P. Let $s_1, s_2, \cdots, s_r$ include all of these points and let $P_1(f)=P(\phi_1)f(s_1)+P(\phi_2)f(s_2)+\cdots+P(\phi_r)f(s_r)$ for all f in $\mathfrak{F}$ where ϕ_i is the characteristic function of s_i for $i=1, 2, \cdots, r$. Then let $P_2=P-P_1$. It is obvious that P_2 is linear and vanishes on characteristic functions of points. Finally since any non-negative member of $\mathfrak{F}$ is the sum of a function which vanishes at the s_i and a linear combination of the ϕ_i with non-negative coefficients it is easily verified that P_2 is non-negative.

LEMMA 3. *If P is a non-negative linear functional on $\mathfrak{F}$ which vanishes on all characteristic functions then P vanishes identically.*

PROOF. It is obviously sufficient to show that $P(f)=0$ whenever $f \geqq 0$. For each $i=1, 2, \cdots$ let E_i be the set of all s in S such that $i-1 \leqq f(s) < i$ and let ϕ_i be the characteristic function of E_i. For each s in S there exists one and only one $i=1, 2, \cdots$ such that $s \in E_i$. Let $g(s)=if(s)$. For each $n=1, 2, \cdots$, $g-\phi_1 g-\phi_2 g-\cdots-\phi_{n-1}g \geqq n(f-\phi_1 f-\phi_2 f-\cdots-\phi_{n-1}f)$ and it is obvious that $P(h)=0$ whenever h is a bounded function. Hence for each $n=1, 2, \cdots$, $0 \leqq P(f) \leqq P(g)/n$. Thus $P(f)=0$.

LEMMA 4. *Let $\phi_1, \phi_2, \cdots$ be the characteristic functions of mutually*

disjoint subsets of S and let ϕ be the characteristic function of their union. Then if P is a non-negative linear functional on $\mathfrak{F}$, $P(\phi)=P(\phi_1)+P(\phi_2)+\cdots$.

PROOF. By Lemma 1 there exists n_0 such that for $n>n_0$, $P(\phi_n)=0$. Since $P(\phi-\phi_1-\cdots-\phi_{n_0})=P(\phi)-P(\phi_1)-\cdots-P(\phi_{n_0})$ we have only to show that $P(\phi-\phi_1-\cdots-\phi_{n_0})=0$. In other words we may confine ourselves to the case in which $P(\phi_n)=0$ for $n=1, 2, \cdots$. For each s in S if $\phi_n(s)=1$ for some $n=1, 2, \cdots$ let $g(s)=n$; otherwise let $g(s)=0$. Then for each $n=1, 2, \cdots$, $g-\phi_1-2\phi_2-\cdots-(n-1)\phi_{n-1}\geqq n(\phi-\phi_1-\cdots-\phi_{n-1})$. Hence $P(g)\geqq nP(\phi)$. In other words $0\leqq P(\phi)\leqq P(g)/n$ for $n=1, 2, \cdots$. Thus $P(\phi)=0$.

We may now prove our equivalence theorem.

THEOREM. *Every bounded linear functional on $\mathfrak{F}$ is a finite linear combination of point functionals if and only if there exists no countably additive measure α defined on all subsets of S which is zero at points, which takes on only the values zero and one and which does not vanish identically.*

PROOF. Suppose that a measure of the sort described does exist. Let f be an arbitrary member of $\mathfrak{F}$. Using the fact that α takes on only one nonzero value, it is easy to prove the existence of a sequence $I_1, I_2, \cdots$ of closed intervals on the real line such that for each $n=1, 2, \cdots$, $I_n\supseteq I_{n+1}$, $\alpha(S-f^{-1}(I_n))=0$, and I_n is of length $1/(2^n)$. Let λ be the unique real number contained in all of the I_n's. Then it is clear that $f(s)=\lambda$ except on a set of α measure zero. In other words for each f in $\mathfrak{F}$ there is a unique real number $P(f)$ such that $f(s)=P(f)$ for "almost all" s. It is obvious that P is a non-negative linear functional on $\mathfrak{F}$ which is not identically zero and which does not vanish on characteristic functions of points. It follows then that P is a bounded linear functional on $\mathfrak{F}$ which is not a finite linear combination of point functionals. Conversely suppose that there exists a bounded linear functional on $\mathfrak{F}$ which is not a finite linear combination of point functionals. As we have already remarked we may suppose that this functional is non-negative. Hence by Lemma 2 there exists a non-negative linear functional P on $\mathfrak{F}$ which is not identically zero and which vanishes on characteristic functions of points. For each subset E of S let $\beta(E)=P(\phi_E)$ where ϕ_E is the characteristic function of E. It follows from Lemma 4 that β is a countably additive measure function and from Lemma 3 that β is not identically zero. Furthermore it is an easy consequence of Lemma 1 that $S=S_1\cup S_2\cup\cdots\cup S_n$ where each S_i has the property that for each subset E of S either $\beta(E\cap S_i)=0$ or

$\beta(E\cap S_i)=\beta(S_i)$. For at least one $i_0=1, 2, \cdots, n$, $\beta(S_{i_0})\neq 0$. Let $\alpha(E)=\beta(E\cap S_{i_0})/\beta(S_{i_0})$ for each subset E of S. Then it is easily verified that α is a measure function with all properties listed in the statement of the theorem.

Using Ulam's results on two-valued measures we deduce the following corollary.

COROLLARY. Let $C_0=\aleph_0$ and for each $n=1, 2, \cdots$ let $C_n=2^{C_{n-1}}$. Let $D=C_0+C_1+\cdots$. Then whenever the cardinal of S is less than D every bounded linear functional on $\mathfrak{F}$ is a finite linear combination of point functionals.

REFERENCES

1. G. Birkhoff, *Lattice Theory*, Amer. Math. Soc. Colloquium Publications, vol. 25, New York, 1940.

2. S. Ulam, *Zur Masstheorie in der allgemeinen Mengenlehre*, Fund. Math. vol. 16 (1930) pp. 140–150.

HARVARD UNIVERSITY

THE ROLE OF INTERNAL FAMILIES IN MEASURE THEORY

ANTHONY P. MORSE

1. **Introduction.** Theorem 4.7 below is an abstract formulation of a certain closed subset theorem[1] recently established by Randolph and myself. It has a wider range of application than similar abstractions due to Hahn[2] and to Saks.[3]

2. **Notation and terminology.** When H is a family of sets we agree that

$$\sigma(H) = \sum_{\beta \in H} \beta, \qquad \pi(H) = \prod_{\beta \in H} \beta.$$

A family R is said to be: *finitely additive* if $\sigma(H) \in R$ whenever H is a finite nonvacuous subfamily of R; *countably additive* if $\sigma(H) \in R$ whenever H is a countable nonvacuous subfamily of R; *finitely multiplicative* if $\pi(H) \in R$ whenever H is a finite nonvacuous subfamily of R; *countably multiplicative* if $\pi(F) \in R$ whenever F is a countable nonvacuous subfamily of R; α *complemental* if R is such a family of subsets of α that $\alpha - \beta \in R$ whenever $\beta \in R$.

If R is a family of sets we also agree that: R_σ is the family of all sets of the form $\sigma(H)$ where H is a countable nonvacuous subfamily of R; R_δ is the family of all sets of the form $\pi(H)$ where H is a countable nonvacuous subfamily of R; R_γ is the family of all sets of the form $\sigma(R) - \beta$ where $\beta \in R$; R^γ is the smallest $\sigma(R)$ complemental, countably additive family which contains R; R^δ is the smallest countably multiplicative, countably additive family which contains R.

DEFINITION 2.1. R is *internal* if and only if R_δ is finitely additive and $R_\gamma \subset R^\delta$.

REMARK 2.2 If R is the family of all closed subsets of a metric space then R is internal[4] and the members of R^γ are the Borel subsets of the space.

Received by the editors November 15, 1943.

[1] A. P. Morse and J. F. Randolph, *The ϕ rectifiable subsets of the plane*, Trans. Amer. Math. Soc. vol. 55 (1944) pp. 236–305, Theorem 3.7 together with the remarks which follow Theorem 3.4.

[2] H. Hahn, *Über die Multiplikation total-additiver Mengenfunktionen*, Annali della R. Scuola Normale Superiore Pisa (2) vol. 2 (1933) p. 437.

[3] S. Saks, *Theory of the integral*, Warsaw, 1937, p. 85.

[4] Since an open set is an R_σ.

3. Two known results in set theory.

THEOREM 3.1. *R_δ is countably multiplicative. If R is finitely additive then so is R_δ.*

PROOF. R_δ is clearly countably multiplicative. The remainder of the theorem follows from the identity

$$\prod_{x\in A}\sum x+\prod_{y\in B}\sum y=\prod_{x\in A}\prod_{y\in B}\sum(x+y).$$

THEOREM 3.2.[5] *If $R_\gamma\subset R^\delta$ then $R^\gamma=R^\delta$.*

PROOF. Let $\tilde{\alpha}=\sigma(R)-\alpha$. Let

$$P=\underset{\alpha}{E}\,[(\alpha\in R^\delta)(\tilde{\alpha}\in R^\delta)].$$

A simple check reveals that P is a $\sigma(R)$ complemental, countably additive subfamily of R^δ. Our assumption that R_γ is contained in R^δ assures us, on the other hand, that P contains R. Accordingly $R^\gamma\subset P\subset R^\delta$. Now R^γ, being $\sigma(R)$ complemental and countably additive, is clearly countably multiplicative also. Consequently $R^\delta\subset R^\gamma$ and the desired conclusion is at hand.

4. The role of internal families in measure theory.

DEFINITION 4.1. We say ϕ *measures* S if and only if ϕ is such a function on $E_\beta[\beta\subset S]$ to $E_t[0\leqq t\leqq\infty]$ that:

I. $\phi(0)=0$;

II. $\phi(A)\leqq\phi(B)$ whenever $A\subset B\subset S$;

III. If H is any countable family for which $\sigma(H)\subset S$, then

$$\phi[\sigma(H)]\leqq\sum_{\beta\in H}\phi(\beta).$$

THEOREM 4.2. *If ϕ measures S and ϕ measures T then $S=T$.*

Due to Carathéodory[6] is

DEFINITION 4.3. A set A is ϕ *measurable* if and only if ϕ measures some superset S of A in such a way that

$$\phi(T)=\phi(TA)+\phi(T-A)$$

whenever $T\subset S$.

[5] This is a corollary of a theorem proved by W. Sierpinski in his *Les ensembles boreliens abstraits*, Annales de la Société polonaise de mathématique vol. 6 (1927) p. 51.

[6] C. Carathéodory, *Über das lineare mass von Punktmengen*, Nachr. Ges. Wiss. Göttingen (1914) p. 406.

THEOREM 4.4. *If R is a family of ϕ measurable sets, ϕ measures $\sigma(R)$, then R^3 and R^7 are families of ϕ measurable sets.*

PROOF. Let M be the family of all ϕ measurable sets. Clearly M is $\sigma(R)$ complemental and countably additive.[7] Consequently $R^3 \subset R^7 \subset M$.

LEMMA 4.5. *If R_3 is a finitely additive family of ϕ measurable sets, ϕ measures $\sigma(R)$, $\phi[\sigma(R)] < \infty$, $B \in R^3$, $\epsilon > 0$, then B contains such a member C of R_3 that $\phi(B - C) < \epsilon$.*

PROOF. Let K be so defined that $\beta \in K$ if and only if corresponding to each positive number η there is such a member C of R_3 that

$$C \subset \beta, \qquad \phi(\beta - C) < \eta.$$

We shall complete the proof by showing in Part III below that $B \in K$.

Part I. *If H is a countable nonvacuous subfamily of K then $\sigma(H) \in K$ and $\pi(H) \in K$.*

PROOF. Let $\eta > 0$. Let $A_1, A_2, A_3, \cdots$ be a sequence whose range is H. Let $C_1, C_2, C_3, \cdots$ be such members of R_3 that

$$C_n \subset A_n, \qquad \phi(A_n - C_n) < \frac{\eta}{2^n}$$

for each positive integer n.

Now

$$\phi\left[\sigma(H) - \sum_{n=1}^{\infty} C_n\right] = \phi\left[\sum_{n=1}^{\infty} A_n - \sum_{n=1}^{\infty} C_n\right] \leqq \phi\left[\sum_{n=1}^{\infty} (A_n - C_n)\right]$$

$$\leqq \sum_{n=1}^{\infty} \phi(A_n - C_n) < \sum_{n=1}^{\infty} \frac{\eta}{2^n} = \eta.$$

Accordingly if N is a sufficiently large integer we are sure that

$$\sum_{n=1}^{N} C_n \in R_3, \qquad \sum_{n=1}^{N} C_n \subset \sigma(H), \qquad \phi\left[\sigma(H) - \sum_{n=1}^{N} C_n\right] < \eta.$$

On the other hand

$$\pi(H) = \prod_{n=1}^{\infty} A_n$$

---

[7] Those measure theoretic results of which we assume a previous knowledge are in H. Hahn, *Theorie der reellen Funktionen*, vol. 1, Berlin, 1921, pp. 424–427.

and $\prod_{n=1}^{\infty} C_n$ is such a member (see 3.1) of R_δ that

$$\prod_{n=1}^{\infty} C_n \subset \pi(H),$$

$$\phi\left[\pi(H) - \prod_{n=1}^{\infty} C_n\right] = \phi\left[\sum_{n=1}^{\infty} \{\pi(H) - C_n\}\right] \leqq \phi\left[\sum_{n=1}^{\infty} (A_n - C_n)\right]$$

$$\leqq \sum_{n=1}^{\infty} \phi(A_n - C_n) < \sum_{n=1}^{\infty} \frac{\eta}{2^n} = \eta.$$

Part II. $R \subset K$.

PROOF. $R \subset R_\delta \subset K$.

Part III. $B \in K$.

PROOF. Parts I and II assure us that K is a countably multiplicative, countably additive family which contains R. Consequently $R^\delta \subset K$ and the conclusion that $B \in K$ follows from our hypothesis that $B \in R^\delta$.

THEOREM 4.6. *If R_δ is a finitely additive family of ϕ measurable sets, ϕ measures $\sigma(R)$, $B \in R^\delta$, $\phi(B) < \infty$, $\epsilon > 0$, then B contains such a member C of R_δ that $\phi(B - C) < \epsilon$.*

PROOF. Let Φ be such a function on the subsets of $\sigma(R)$ that

$$\Phi(\alpha) = \phi(B\alpha) \quad \text{whenever} \quad \alpha \subset \sigma(R).$$

Check that Φ measures $\sigma(R)$ and that 4.5 may be applied to yield the desired conclusion.

THEOREM 4.7. *If R is an internal family of ϕ measurable sets, ϕ measures $\sigma(R)$, $B \in R^\gamma$, $\phi(B) < \infty$, $\epsilon > 0$, then B contains such a member C of R_δ that $\phi(B - C) < \epsilon$.*

PROOF. Use 4.6, 2.1, and 3.2.

DEFINITION 4.8. We say ϕ is a *Borelian measure* with respect to R if and only if: R is an internal family of ϕ measurable sets; ϕ measures $\sigma(R)$; corresponding to each subset A of $\sigma(R)$ there is a set β for which

$$\beta \in R^\gamma, \qquad A \subset \beta, \qquad \phi(A) = \phi(\beta).$$

THEOREM 4.9. *If ϕ is a Borelian measure with respect to R, A is a ϕ measurable set, $\phi(A) < \infty$, $\epsilon > 0$, then A contains such a member C of R_δ that $\phi(A - C) < \epsilon$.*

PROOF. Let B', B'', B''' be such sets that

$$A \subset B' \in R^\gamma, \qquad \phi(B') = \phi(A),$$

$$B' - A \subset B'' \in R^\gamma, \qquad \phi(B'') = \phi(B' - A),$$
$$B''' = B' - B''.$$

Clearly

$$B''' \in R^\gamma, \qquad B''' = B' - B'' \subset B' - (B' - A) \subset A,$$
$$\phi(A - B''') \leqq \phi(B' - B''') \leqq \phi(B'') = \phi(B') - \phi(A) = 0.$$

Application of 4.7 to the set B''' completes the proof.

THEOREM 4.10. *If R is the family of all closed subsets of a metric space S, ϕ measures S in such a way that closed sets are ϕ measurable, B is a Borel set, $\phi(B) < \infty$, $\epsilon > 0$, then B contains such a closed set C that $\phi(B - C) < \epsilon$.*

PROOF. Clearly R is an internal family for which $R = R_\delta$, and $\sigma(R) = S$. Application of 4.7 completes the proof. Using 4.9 we obtain

THEOREM 4.11. *If R is the family of all closed subsets of a metric space S, ϕ is a Borelian measure with respect to R, A is ϕ measurable, $\phi(A) < \infty$, $\epsilon > 0$, then A contains such a closed set C that $\phi(A - C) < \epsilon$.*

REMARK 4.12. Theorems 4.9 and 4.11 are generalizations of a result due to Hahn.[8] For corollaries and special cases of Theorems 4.7, 4.9, 4.10, and 4.11, see Saks, op. cit., Theorem 6.5 on page 68, Theorem 6.6 on page 69, the correct portions of Theorem 9.7+ on page 85, the proof of Lemma 5.1 on page 114, Lemma 15.1 on page 152.

Let us now examine, in the light of an example, the just cited Theorem 9.7+ and my own Theorem 4.7. Let S be the ordinary real numbers metrized in the customary manner. Let F be the family of all closed subsets of S, G the family of all open subsets of S. Let $R = F_\sigma G_\delta$. It is easily seen, with the aid of 3.1, that R is a finitely additive, S complemental, internal family. Furthermore $\sigma(R) = S$ and R^γ is precisely the family of all Borel subsets of S. Let B be the rational numbers and let ϕ so measure S that

$$\phi(\beta) = \text{the number of numbers in } \beta B$$

whenever $\beta \subset S$. Note that $\phi(B) = \phi(S) = \infty$ but that S is a countable sum of Borel sets of finite ϕ measure. However, within the Borel set B, it is impossible to find a G_δ set C for which $\phi(B - C) < 1$; if this could be done then C would equal B and B itself would be a G_δ in contradiction to the well known fact that a dense G_δ is a residual set with the power of the continuum. Since $R_\delta \subset G_\delta$ it is also impossible to find, within the Borel set B, an R_δ set C for which $\phi(B - C) < 1$.

[8] H. Hahn, *Theorie der reellen Funktionen*, vol. 1, Berlin, 1921, p. 447, Theorem IV.

This reveals the essential nature of the hypothesis "$\phi(B) < \infty$" in 4.7 as well as the erroneous aspects of the "more generally" part of Saks' Theorem 9.7+. Nevertheless it is easy to verify the statement obtained from Theorem 4.10 by deleting the hypothesis "$\phi(B) < \infty$" and replacing it by "each bounded set has finite ϕ measure."

REMARK 4.13. Herein we give a supplementary example which serves much the same purpose as the one just discussed in 4.12. Let S be the plane metrized in the customary manner. Introduce F, G, and R as in 4.12. Let B be those points in the plane whose first coordinates are rational. Let ϕ so measure S that

$$\phi(\beta) = \text{the Carathéodory}^9 \text{ linear measure of } \beta B$$

whenever $\beta \subset S$. Note that $\phi(B) = \phi(S) = \infty$ but that S is a countable sum of Borel sets of finite ϕ measure. Note also (cf. 4.12) that each countable subset of S has ϕ measure zero. However, within the Borel set B, it is impossible to find a G_δ set C for which $\phi(B - C) < \infty$. To see this use the fact that the projection upon the y axis of any subset α of B has a Lebesgue measure which does not exceed $\phi(\alpha)$, and then recall the reasoning employed in 4.12.

UNIVERSITY OF CALIFORNIA

[9] C. Carathéodory, op. cit., pp. 420 ff.

ON LINEAR EQUATIONS IN HILBERT SPACE

L. W. COHEN

Given an infinite matrix $A=\|a_{ij}\|$ where a_{ij} is complex and

$$\sum_{j=1}^{\infty}|a_{ij}|^2<+\infty, \qquad i=1,2,\cdots, \tag{1}$$

the problem of solving the system of linear equations

$$y_i=\sum_{j=1}^{\infty}a_{ij}x_j, \qquad i=1,2,\cdots, \tag{2}$$

has been studied from several points of view. For arbitrary y_i, E. Schmidt[1] has given necessary and sufficient conditions on the a_{ij}, y_i so that the system (2) have a solution $x=(x_j)\in H_2$ (Hilbert space). Schmidt shows that if a solution exists, the solution of minimum norm is unique, and gives explicit formulas for this solution. If A defines a linear transformation T on H_2 to H_2, F. Riesz[2] gives necessary and sufficient conditions that an inverse T^{-1} exist, that is, that the solution $x=T^{-1}(y)$ where T^{-1} is a linear transformation. The following problem stands between these two: Find conditions on the elements of A so that the system (2) have a solution $x\in H_2$ for each $y\in H_2$. Such conditions will permit the use of Schmidt's formulas to express the minimal solution x for each y but this of course does not imply the existence of an inverse of the matrix A. We give a solution of this problem by a method which depends on a property, which seems new, of the m-rowed minors of the matrices $A_{i_1\cdots i_m}=\|a_{i_s,j}\|_{1\le s\le m;\,j\ge 1}$ and on Cramer's rule.

Let

$$a(i_1,\cdots,i_m;j_1,\cdots,j_m)=\det\|a_{i_s,j_t}\|_{1\le s,t\le m}$$

be the determinant of the columns $j_1,\cdots,j_m$ of $A_{i_1\cdots i_m}$. If $B=\|b_{ij}\|$ satisfies (1) and $B'_{i_1\cdots i_m}$ is the transposed of $B_{i_1\cdots i_m}$, the determinant $\det A_{i_1\cdots i_m}B'_{i_1\cdots i_m}=\det\|\sum_{k=1}^{\infty}a_{i_s,k}b_{i_t,k}\|_{1\le s,t\le m}$ is finite. Because of the continuity of a determinant as a function of its elements

$$\det A_{i_1\cdots i_m}B'_{i_1\cdots i_m}=\lim_{n}\det\left\|\sum_{k=1}^{n}a_{i_s,k}b_{i_t,k}\right\|_{1\le s,t\le m}. \tag{3}$$

Presented to the Society April 23, 1943; received by the editors March 24, 1943.

[1] E. Schmidt, *Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten*, Rend. Circ. Mat. Palermo vol. 25 (1908) pp. 53–77.

[2] F. Riesz, *Les systèmes d'equations linéaires a une infinité d'inconnues*, Paris, 1913, p. 86.

There is a theorem[3] on the minors of products of square matrices which, with slight modification in its proof, yields the identity

$$\det \left\| \sum_{k=1}^{n} a_{i_s k} b_{i_t k} \right\|_{1 \leq s,t \leq m} = \sum_{[i_1 \cdots i_m]}^{n} a(i_1, \cdots, i_m; j_1, \cdots, j_m) b(i_1, \cdots, i_m; j_1, \cdots, j_m), \quad n \geqq m, \tag{4}$$

where the sum is extended over all combinations $j_1, \cdots, j_m$ in $1, \cdots, n$.

THEOREM 1. *If A, B satisfy* (1), *then*

$$\det A_{i_1 \cdots i_m} B'_{i_1 \cdots i_m} = \sum_{[i_1 \cdots i_m]}^{\infty} a(i_1, \cdots, i_m; j_1, \cdots, j_m) b(i_1, \cdots, i_m; j_1, \cdots, j_m),$$

where the sum is extended over all combinations of positive integers $j_1, \cdots, j_m$. The series converges absolutely and

$$|\det A_{i_1 \cdots i_m} B'_{i_1 \cdots i_m}| \leqq [\det A_{i_1 \cdots i_m} \overline{A}'_{i_1 \cdots i_m}]^{1/2} [\det B_{i_1 \cdots i_m} \overline{B}'_{i_1 \cdots i_m}]^{1/2}.$$

PROOF. By Schwartz' inequality we have, from (4) and (3) with $B = \overline{A}$,

$$\begin{aligned} \sum_{[i_1 \cdots i_m]}^{n} & |a(i_1, \cdots, i_m; j_1, \cdots, j_m) b(i_1, \cdots, i_m; j_1, \cdots, j_m)| \\ & \leqq \left[\sum_{[i_1 \cdots i_m]}^{n} |a(i_1, \cdots, i_m; j_1, \cdots, j_m)|^2 \right]^{1/2} \cdot \left[\sum_{[i_1 \cdots i_m]}^{n} |b(i_1, \cdots, i_m; j_1, \cdots, j_m)|^2 \right]^{1/2} \\ & \leqq [\det A_{i_1 \cdots i_m} \overline{A}'_{i_1 \cdots i_m}]^{1/2} [\det B_{i_1 \cdots i_m} \overline{B}'_{i_1 \cdots i_m}]^{1/2}, \qquad n \geqq m. \end{aligned}$$

The conclusion is now evident as a consequence of (3).

If we define

$$|AB'| = \limsup_{m} \sup_{i_1 \cdots i_m} \left| \sum_{[i_1 \cdots i_m]}^{\infty} a(i_1, \cdots, i_m; j_1, \cdots, j_m) x \times b(i_1, \cdots, i_m; j_1, \cdots, j_m) \right|$$

[3] C. C. MacDuffee, *An introduction to abstract algebra*, New York, 1940, Theorem 99.1, p. 216.

we have a Schwarz inequality for matrices:

$$|AB'| \leqq |A\overline{A}'|^{1/2} |B\overline{B}'|^{1/2}.$$

The following lemma contains the Gram condition for linear dependence.

LEMMA 1. *The following statements are equivalent:*
(a) *The rows of* $A_{1 \cdots m}$ *are linearly dependent.*
(b) Det $A_{1 \cdots m}\overline{A}'_{1 \cdots m} = 0$.
(c) *All* m*-rowed minors of* $A_{1 \cdots m}$ *equal zero.*

PROOF. That (a) implies (c) is immediate. The equivalence of (b) and (c) follows from Theorem 1 with $A = B$. It remains to show that (c) implies (a). This is evident if $m = 1$. Assuming this statement for $m-1$, it is true for m if all the $(m-1)$-rowed minors of A_{m-1} vanish. If one such minor does not vanish, say

$$\det \|a_{ij}\|_{1 \leqq i, j \leqq m-1} \neq 0,$$

we denote by c_k the cofactor of a_{km} in the determinant of the first m columns of $A_{1 \cdots m}$. Then $c_m \neq 0$ and

$$\sum_{k=1}^{m} c_k a_{kj} = 0, \qquad j = 1, \cdots, m-1.$$

But this sum vanishes for all other values of j because of (c). Hence (c) implies (a).

THEOREM 2. *If* A *satisfies* (1), *the finite system*

$$y_i = \sum_{j=1}^{\infty} a_{ij} x_j, \qquad i = 1, \cdots, m, \tag{5}$$

has a solution $x \in H_s$ *for each* $y_1, \cdots, y_m$ *if and only if*

$$\det A_{1 \cdots m}\overline{A}'_{1 \cdots m} \neq 0.$$

PROOF. The necessity is a consequence of Lemma 1. If the condition is satisfied, then there is a nonvanishing $a(1, \cdots, m; j_1, \cdots, j_m)$ by (c) of Lemma 1 and a solution $x = (x_j)$ where $x_j = 0$ for $j \neq j_1, \cdots j_m$ and $x_{j_1}, \cdots, x_{j_m}$ are determined by Cramer's rule.

COROLLARY. *If* A *satisfies* (1) *and the system* (2) *has a solution* $x \in H_2$ *for each* $y \in H_2$, *then* det $A_{1 \cdots m}\overline{A}'_{1 \cdots m} \neq 0$ *for all* m.

An estimate of the minimum norm of the solution of the finite system (5) may be given in terms of a series of finite minors in $A_{i_1 \cdots i_m}$.

Let $J=[j_1, \cdots, j_m]$ be a combination of m distinct positive integers and let S_m be a set of J such that no two J's have a common integer while every positive integer is in some $J \in S_m$. Let

$$\alpha_{i_1 \cdots i_m} = \left[\sup_{S_m} \sum_{J \in S_m} | a(i_1, \cdots, i_m; j_1, \cdots, j_m) |^2 \right]^{1/2},$$

$$a_{i_1 \cdots i_m} = \left[\sum_{[j_1 \cdots j_m]} | a(i_1, \cdots, i_m; j_1, \cdots, j_m) |^2 \right]^{1/2}.$$

Since S_m is a subset of the sum of all $j_1, \cdots, j_m$ we have the following lemma.

LEMMA 2. $\alpha_{i_1 \cdots i_m} \leqq a_{i_1 \cdots i_m}$,

$$\sup_{S_m} \sum_{J \in S_m} \sum_{s=1}^{m} |a(1, \cdots, k-1, k+1, \cdots, m; j_1, \cdots, j_{s-1} j_{s+1}, \cdots, j_m)|^2$$

$$\leqq \sum_{[j_1 \cdots j_{m-1}]}^{\infty} | a(1, \cdots, k-1, k+1, \cdots, m; j_1, \cdots, j_{m-1}) |^2.$$

THEOREM 3. *If A satisfies* (1) *and the finite system* (5) *has a solution $x^m \in H_2$ for each $y^m = (y_1, \cdots, y_m, 0, 0, \cdots)$, then*

$$\inf \|x^m\| \leqq \|y^m\| \left[\sum_{k=1}^{m} \left| \frac{a_{1, \cdots, k-1, k+1, \cdots, m}}{\alpha_{1 \cdots m}} \right|^2 \right]^{1/2}.$$

PROOF. By Theorem 2, $\det A_1 \cdots {}_m \overline{A}_1' \cdots {}_m \neq 0$ and so $\alpha_{1 \cdots m} \neq 0$. Let $M^m_{kj_s}$ be the cofactor of a_{kj_s} in $a(1, \cdots, m; j_1, \cdots, j_m)$. The system (5) has a solution $x^m_{j_1 \cdots j_m}$ defined by

$$x^m_{j; j_1 \cdots j_m} = \begin{cases} \sum_{k=1}^{m} y_k M^m_{kj} / a(1, \cdots, m; j_1, \cdots, j_m), & j = j_1, \cdots, j_m, \\ 0, & j \neq j_1, \cdots, j_m. \end{cases}$$

We have

$$\|x^m_{j_1 \cdots j_m}\|^2 \, | \, a(i_1, \cdots, i_m; j_1 \cdots j_m) \, |^2$$

$$\leqq \sum_{s=1}^{m} \left| \sum_{k=1}^{m} y_k M^m_{kj_s} \right|^2 \leqq \|y^m\|^2 \sum_{s=1}^{m} \sum_{k=1}^{m} | M^m_{kj_s} |^2$$

$$= \|y^m\|^2 \sum_{s=1}^{m} \sum_{k=1}^{m} | a(1, \cdots, k-1, k+1, \cdots, m; j_1, \cdots, j_{s-1}, j_{s+1}, \cdots, j_m) |^2.$$

Hence

$$\inf \|x^m\|^2 \alpha^2_{1\cdots m} \leqq \|y^m\|^2 \sum_{k=1}^{m} \sup_{\delta_m} \sum_{J\in\delta_m} \sum_{s=1}^{m} |a(1,\cdots,k-1,k+1,\cdots,m;j_1\cdots j_{s-1}j_{s+1}\cdots j_m)|^2$$
$$\leqq \|y^m\|^2 \sum_{k=1}^{m} a^2_{1,\cdots,k-1,k+1,\cdots,m}$$

by Lemma 2. The conclusion follows at once.

A sufficient condition for the solution of the system (2) for each $y\in H$ may be obtained by restricting the constants

$$\alpha_m = \left[\sum_{k=1}^{m} \left|\frac{a_{1\cdots k-1,k+1,\cdots,m}}{\alpha_{1\cdots m}}\right|^2\right]^{1/2}, \qquad m = 1, 2, \cdots.$$

THEOREM 4. *If A satisfies* (1), *its rows are linearly independent, and* $\alpha = \liminf_m \alpha_m < +\infty$, *then for each* $y\in H_2$ *the system* (2) *has a solution* $x\in H_2$ *such that*

$$\|x\| \leqq \alpha\|y\|.$$

PROOF. Consider any $y\in H_2$ and any $\epsilon>0$. The sequence contains a subsequence $\alpha_{m_\mu}<\alpha+\epsilon$. From Theorem 3 it follows that for each μ there is an $x^\mu=(x^\mu_j)\in H_2$ such that

$$y_i = \sum_{j=1}^{\infty} a_{ij}x^\mu_j, \qquad i = 1, \cdots, m_\mu,$$
$$\|x^\mu\| \leqq (\alpha+\epsilon)\|y\|.$$

Applying a diagonal process to (x^μ_j) one finds a subsequence $x^{\mu_\nu}=(x^{\mu_\nu}_j)$ and an $x=(x_j)\in H_2$ such that

$$\lim_\nu x^{\mu_\nu}_j = x_j, \qquad j = 1, 2, \cdots,$$
$$\|x\| \leqq (\alpha+\epsilon)\|y\|.$$

Since for all ν, $N>0$ and $1\leqq i\leqq m_{\mu_\nu}$

$$y_i - \sum_{j=1}^{\infty} a_{ij}x_j = \sum_{j=1}^{N} a_{ij}(x^{\mu_\nu}_j - x_j) + \sum_{j=N+1}^{\infty} a_{ij}x^{\mu_\nu}_j - \sum_{j=N+1}^{\infty} a_{ij}x_j,$$

x solves the system (1).

Now consider $\epsilon_n\downarrow 0$. For each n there is an $x^n\in H_2$ which solves the system (2) and such that $\|x^n\|\leqq(\alpha+\epsilon_n)\|y\|$. Repeating the diagonal process and the above argument, one finds an $x\in H_2$ such that $\|x\|\leqq\alpha\|y\|$ and which solves the system (2).

UNIVERSITY OF WISCONSIN

A NOTE ON AN INEQUALITY OF E. SCHMIDT

RICHARD BELLMAN

In a note inserted in the Sitzungsberichte der Preussicher Akademie, 1932, E. Schmidt stated without proof a relation

$$\left[\int_{-1}^{1} f'(x)^2 dx\right]^{1/2} \leqq k_N N^2 \left[\int_{-1}^{1} f(x)^2 dx\right]^{1/2} \tag{1}$$

where $f(x)$ is a polynomial of degree N, and it was subsequently shown by Hille, Szego and Tamarkin that k_N is a bounded function of N such that $\lim_{N\to\infty} k_N = \pi^{-1}$.

In a joint paper, generalizing the classical Markoff theorem

$$\max_{-1\leq x\leq 1} |f'(x)| \leqq N^2 \max_{-1\leq x\leq 1} |f(x)|$$

to general mean values of the form $\int_{-1}^{1}|f(x)|^p dx$, $p \geqq 1$, Hille, Szego and Tamarkin [1][1] gave three proofs of the inequality (1), two products of the general case of exponent p, and another for the case $p=2$, stated by the authors to be similar to the original unpublished proof of Schmidt.

It seems worthwhile to sketch a short elementary proof of the important case $p=2$, depending only upon an elementary inequality:[2]

$$n\sum_{1}^{n} a_k^2 - \left(\sum_{1}^{n} a_k\right)^2 \geqq 0 \tag{2}$$

and a simple property of the Legendre polynomials.

The required result for Legendre polynomials is [2] $P'_{n+1}(z) - P_{n-1}(z) = (2n+1)P_n(z)$, where $P_n(z)$ is the nth Legendre polynomial. From this we obtain

$$\begin{aligned} P'_{2n} &= \sum_{0}^{n-1} (4k+3)P_{2k+1}, \\ P'_{2n+1} &= \sum_{0}^{n} (4k+1)P_{2k}. \end{aligned} \tag{3}$$

Let the polynomial of the Nth degree be expressed in Legendre polynomials

Received by the editors January 31, 1944.

[1] Numbers in brackets refer to the references cited at the end of the paper.

[2] The use of this inequality, instead of a more complicated one due to Hardy, was suggested by Professor Szász. It simultaneously simplifies the proof and yields a better constant than that originally obtained by the author.

$$f_N(x) = \sum_{k=0}^{N} a_k P_k(x).$$

Then, using the well known orthogonality relations of the P_n,

$$\int_{-1}^{1} f_N(x)^2 dx = 2\sum_{0}^{N} \frac{a_k^2}{2k+1}.$$

Using (3),

$$f_N'(x) = \sum_{2k \leqq N-1} (4k+1) P_{2k}(x) \Big[\sum_{n \geqq k+1} a_{2n+1} \Big]$$

$$+ \sum_{2k \leqq N-1} (4k+3) P_{2k+1}(x) \Big[\sum_{n \geqq k+1} a_{2n} \Big].$$

Thus

$$\int_{-1}^{1} f_N'(x)^2 dx = 2 \sum_{2k \leqq N-1} (4k+1) \Big[\sum_{r \geqq k} a_{2r+1} \Big]^2$$

$$+ 2 \sum_{2k \leqq N-1} (4k+3) \Big[\sum_{r \geqq k} a_{2n} \Big]^2.$$

Let $a_r = ((2r+1)/2)^{1/2} b_r$. Then the first sum is

$$2 \sum_{2k \leqq N-1} (4k+1) \Big[\sum_{r \geqq k} a_{2r+1} \Big]^2$$

$$\leqq 2 \Big[\sum_{2k \leqq N-1} (4k+1) \Big] \Big[\sum \Big(\frac{4k+1}{2} \Big)^{1/2} b_{2k+1} \Big]^2$$

$$\leqq \frac{N(N+1)(2N+1)}{4} [\sum b_{2k+1}]^2$$

$$\leqq (N+1)^4/2 [\sum b_{2k+1}^2]$$

upon using the inequality (2) upon $[\sum b_{2k+1}]^2$.

The same process yields the same estimate for the second sum. Thus

$$\int_{-1}^{1} f_N'(x)^2 dx \leqq \frac{(N+1)^4}{2} \sum_{0}^{N} b_k^2 = \frac{(N+1)^4}{2} \sum_{0}^{N} \frac{2a_k^2}{2k+1}$$

$$= \frac{(N+1)^4}{2} \int_{-1}^{1} f_N(x)^2 dx.$$

This finishes the proof of (1) and yields lim sup $k_N \leqq 1/2^{1/2}$.

It is clear from the proof that by using Laguerre and Hermite polynomials with the corresponding weight functions, one can obtain similar results for the intervals $(0, \infty)$ and $(-\infty, \infty)$.

References

1. E. Hille, G. Szego, A. Tamarkin, *On some generalizations of a theorem of A. Markoff*, Duke Math. J. vol. 3 (1937) p. 729.
2. Whittaker-Watson, *Modern analysis*, Cambridge, 1935, p. 307 (IV).

Princeton University

TRANSCENDENCE OF CERTAIN CONTINUED FRACTIONS

G. CUTHBERT WEBBER

Even today very little is known about which regular continued fraction expansions represent transcendental numbers. In 1851 Liouville[1] proved the transcendence of the numbers named after him, in the expansions of which the partial quotients increase very rapidly; in 1906 Maillet[2] proved the transcendence of certain expansions in which the partial quotients may be bounded. C. L. Siegel[3] proved that if the partial quotients of the expansion form an arithmetic progression of order one, the number represented is transcendental. In this paper well known transformations of continued fractions are applied, in conjunction with the Siegel theorem, to prove the transcendence of certain classes of continued fraction expansions.

We shall use the notation

$$b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots$$

to represent a continued fraction expansion, with the partial numerators and denominators a_j and b_j respectively. The jth convergent is denoted by A_j/B_j. When the expansion is regular the following notations are also used:

Received by the editors April 6, 1944.

[1] J. Liouville, *Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même reductible à des irrationnelles algébrique*, J. Math. Pures Appl. vol. 16 (1851).

[2] E. Maillet, *Introduction à la théorie des nombres transcendants*, Paris, 1906.

[3] C. L. Siegel, *Über einige Anwendungen diophantischer Approximationen*, Abhandlung der Preussischen Akademie der Wissenschaften, Physikalische-Mathematische Klasse, no. 1 (1929) p. 29.

$$[b_0, b_1, b_2, \cdots] = [b_j]_{j=0}^{\infty}.$$

In the expansions considered $a_j = \pm 1$ $(j \geqq 1)$, the partial quotients are integers, $b_j \geqq 1$ $(j \geqq 1)$, $b_j + a_{j+1} \geqq 2$ for an infinite set of values of j, and $b_j + a_{j+1} \geqq 1$ for all j except, possibly, a finite set. Thus the expansions are semi-regular, at least after a finite number of terms, and so they converge.[4] The expansions considered in the theorems are of Hurwitz type except when one or more partial numerators are -1 in Theorem I.

Siegel, in the paper referred to above, states that the methods applied therein suffice to prove that if $x \neq 0$ is algebraic, then there does not exist between $J_\lambda(x)$ and $J_{\lambda-1}(x)$ a rational equation with rational coefficients, if λ is rational but not the half of an odd integer ($J_\lambda(x)$ is a Bessel function of type one); also, that the same result holds when λ is half an odd integer, due to the special form of $J_\lambda(x)/J_{\lambda-1}(x)$ in terms of e^{ix} and to the Lindemann theorem: The form $\sum_{\nu=1}^{n} \xi_\nu e^{\eta_\nu}$ is transcendental in each non-trivial case, if $\xi_1, \cdots, \xi_n$ and $\eta_1, \cdots, \eta_n$ are algebraic numbers.

The theory of Bessel functions gives the result that

$$\lambda x + \cfrac{a_1}{(\lambda+1)x + \cfrac{a_2}{(\lambda+2)x + \cdots}}$$

is equal to $iJ_{\lambda-1}(2i/x)/J_\lambda(2i/x)$ or $J_{\lambda-1}(2/x)/J_\lambda(2/x)$, according as the a_j are all $+1$ or all -1. If x and λ are chosen such that $x = d$ and $\lambda d = r$, r and d being positive integers, $r \geqq 2$, then

$$r + \cfrac{a_1}{(r+d) + \cfrac{a_2}{(r+2d) + \cdots}}, \tag{1}$$

where the a_j are all $+1$ or all -1, is transcendental.[5]

THEOREM I. *The number ξ, represented by*

$$\xi = g_0 + \cfrac{a_1}{b_1 + \cdots + \cfrac{a_{k-1}}{b_{k-1} + \cfrac{f_1}{g_1 + \cfrac{a_1}{b_1 + \cdots + \cfrac{a_{k-1}}{b_{k-1} + \cfrac{f_2}{g_2 + \cdots}}}}}} \tag{2}$$

is transcendental when the sequence $\{g_j\}$ is an arithmetic progression of first order whose elements are integers which are positive except, possibly, g_0; $b_1, \cdots, b_k$ are positive integers, $a_m = \pm 1$ $(m = 1, \cdots, k-1)$,

[4] O. Perron, *Die Lehre von den Kettenbrüchen*, B. G. Teubner, Leipzig, 1929, pp. 148–154.

[5] If the a_j are all $+1$, this becomes the Siegel theorem stated in the introduction.

the f_j are either all $+1$ or all -1, and $g_1+a_1 \geqq 1$, $b_m+a_{m+1} \geqq 1$, $b_{k-1}+f_j \geqq 1$, for $m=1, \cdots, k-2$ and $j \geqq 1$.

In the proof we may consider $g_0 \geqq 2$; otherwise, the first $k+1$ or $2k+2$ terms are dropped and the remaining terms renumbered to insure convergence of the semi-regular expansion. This does not affect the algebraic character of the whole expansion. The continued fraction (2) is transformed by "contraction"[6] to the continued fraction with convergents A_{jk-1}/B_{jk-1} $(j=1, 2, 3, \cdots)$. This gives

$$\xi = \frac{1}{B_{k-1}}\left\{g_0B_{k-1} + a_1B_{k-2,1} + \frac{(-1)^{k-1}a_1a_2\cdots a_{k-1}f_1}{(g_1B_{k-1} + f_1B_{k-2} + a_1B_{k-2,1})} + \frac{(-1)^{k-1}a_1\cdots a_{k-1}f_2}{(g_2B_{k-1} + f_2B_{k-2} + a_1B_{k-2,1}) + \cdots}\right\}$$

$$= \frac{1}{B_{k-1}}\left\{g_0B_{k-1} + a_1B_{k-2,1} + \frac{(-1)^{k-1}a_1a_2\cdots a_{k-1}f_1}{\eta}\right\},$$[7]

where η is a continued fraction of type (1), at least after a finite number of terms. The latter case may occur when either or both of a_1 and f_1 are -1. Since η is transcendental it follows that ξ is also transcendental.

It is interesting to note that if the expansion is regular, the partial quotients may be made to increase as "slowly as desired" by choosing $b_j=1$ $(j=1, 2, \cdots, k)$ and k large.

THEOREM II. *If a, d and c are positive integers, either c or d being even, and $b=c^2a+(1/2)c^2d$, $e=c^3d$, then*

$$\alpha = [a, b, a+d, b+e, a+2d, b+2e, \cdots]$$

is a transcendental number.

From the theory of continued fractions we obtain the following result:[8]

$$c_0b_0 + \frac{c_0c_1a_1}{c_1b_1 + \cdots + \frac{c_{k-1}c_ka_k}{c_kb_k + \cdots}} = c_0\left[b_0 + \frac{a_1}{b_1 + \cdots + \frac{a_k}{b_k + \cdots}}\right]. \tag{3}$$

[6] O. Perron, *Die Lehre von den Kettenbrüchen*, B. G. Teubner, Leipzig, 1929, pp. 197–203.

[7] By $B_{k-2,1}$ is meant the result of advancing by one the subscripts of all the a_j and b_j in B_{k-2}.

[8] O. Perron, *Die Lehre von den Kettenbrüchen*, B. G. Teubner, Leipzig, 1929, p. 196.

Application of this result with $a_k=1$, $c_k=1/c_{k-1}$, $c_0=c$, $c_{2k}=c$ $(k=1, 2, 3, \cdots)$ to α yields

$$c\alpha = c[a, c^2a + (1/2)c^2d, a + d, c^2a + (3/2)c^2d, a + 2d, c^2a + (5/2)c^2d, \cdots]$$
$$= [ca, ca + (1/2)cd, ca + cd, ca + (3/2)cd, ca + 2cd, ca + (5/2)cd, \cdots].$$

Since the partial quotients in $c\alpha$ form an arithmetic progression of order one with common difference $(1/2)cd$, $c\alpha$, and in turn α, is transcendental by the Siegel theorem stated in the introduction.

In the last theorem we proved the transcendence of a class of regular continued fractions in which the alternating partial quotients form an arithmetic progression of order one. The following theorem gives two more classes of a similar type.

THEOREM III. *The numbers β and γ, where*

$$\beta = [1 + 4j, (3 + 4j)u^2]_{j=0}^{\infty} \text{ and } \gamma = [1, (3 + 4j)u^2 - 2, 1, 3 + 4j]_{j=0}^{\infty},$$

u^2 being a positive integer, are transcendental.

Again appealing to well established theory of continued fractions[9] we have

$$\delta = 0 + \cfrac{z}{1 + \cfrac{a_2z^2}{3 + \cfrac{a_3z^2}{5 + \cfrac{a_4z^2}{7 + \cdots}}}} = \frac{e^z - e^{-z}}{e^z + e^{-z}} \text{ or } \tan z,$$

according as the a_j are all $+1$ or all -1. Apply (3) with $c_0=1/z$, $c_{2j+1}=1$, $c_{2j+2}=1/z^2$ $(j=0, 1, 2, \cdots)$ and $z=1/u$; this yields

$$\delta = z\left[0 + \cfrac{1}{1 + \cfrac{a_2}{(3/z^2) + \cfrac{a_3}{5 + \cfrac{a_4}{(7/z^2) + \cfrac{a_5}{9 + \cdots}}}}}\right]$$
$$= (1/u)\left[0 + \cfrac{1}{1 + \cfrac{a_2}{3u^2 + \cfrac{a_3}{5 + \cfrac{a_4}{7u^2 + \cfrac{a_5}{9 + \cdots}}}}}\right]$$
$$= (1/u)(1/\beta) \quad \text{or} \quad (1/u)(1/\epsilon),$$

according as the a_j are all $+1$ or all -1, where ϵ is defined by the above equality.

Since u^2 is a positive integer, u is an algebraic number not equal to 0, and thus z is an algebraic number. Now by the Lindemann theo-

[9] O. Perron, p. 353.

rem, $\tan z$ and $(e^z-e^{-z})/(e^z+e^{-z})$ are transcendental when $z\neq 0$ is algebraic. This proves the transcendence of β and ϵ.

Convert the semi-regular continued fraction for ϵ to a regular continued fraction;[10] this gives

$$\frac{1}{\epsilon}=\frac{1}{1}-\frac{1}{3u^2}-\frac{1}{5}-\frac{1}{7u^2}-\frac{1}{9}-\cdots,$$

$$\epsilon=0+\frac{1}{1}+\frac{1}{(3u^2-2)}+\frac{1}{1}+\frac{1}{3}+\frac{1}{1}+\frac{1}{(7u^2-2)}+\frac{1}{1}+\frac{1}{7}+\cdots,$$

$$\frac{1}{\epsilon}=[1,(3+4j)u^2-2,1,3+4j]_{j=0}^{\infty}=\gamma.$$

Since ϵ is transcendental, so is γ.

THE UNIVERSITY OF DELAWARE

[10] O. Perron, p. 159.

A NOTE ON A THEOREM OF HARDY ON FOURIER CONSTANTS

RICHARD BELLMAN

In Note LXVI of his *Notes on some points in the integral calculus* [1],[1] Hardy proves the following theorem.

THEOREM 1 (HARDY). *If* $a_1, a_2, \cdots, a_n, \cdots$ *are the Fourier constants of a function of* L^p, $p \geqq 1$, *then* $A_1, A_2, \cdots, A_n, \cdots$ *are also the Fourier constants of a function of* L^p, *where* $A_n = n^{-1}\sum_1^n a_k$.

Hardy restricts the class of functions considered to even functions, with mean value over a period zero, and the same restriction shall be observed here.

We wish to prove a "dual" to Theorem 1, namely:

THEOREM 2. *If* $b_1, b_2, \cdots, b_n, \cdots$ *are the Fourier constants of a function,* $g(x)$, *of* L^p, $p > 1$, *then* $B_1, B_2, \cdots, B_n, \cdots$ *are also the Fourier constants of a function,* $G(x)$, *of* L^p, *where* $B_n = \sum_n^\infty b_k k^{-1}$, *and if* $f(x) \sim \sum a_n \cos nx$ *is any function of* $L^{p'}$, $F(x) \sim \sum_1^\infty A_n \cos nx$, *then,* $1/p + 1/p' = 1$,

$$\int_0^\pi f(x)G(x)dx + \int_0^\pi F(x)g(x)dx = 0.$$

There is a difference between the theorems. The case $p = 1$ is omitted in Theorem 2, and necessarily, since $g(x) \sim \sum_{n=2}^\infty \cos nx/\log n$ belongs to L, but the corresponding B_n do not even exist.

Hardy's method depends upon an explicit representation of the function $F(x)$ in terms of $f(x)$. In the case of the theorem to be proven, this representation does not seem to facilitate matters. The method used will depend upon some general theorems on Fourier series, together with the original theorem of Hardy. For the case $p = 2$, the proof is immediate, since it can be shown that the convergence of b_n^2 implies the convergence of B_n^2 (this result is also due to Hardy, and was the origin of this type of theorem). In particular, this series converges for a function belonging to L^p, for $p \geqq 2$, and thus there exists a function of L^2, having the B_n as Fourier coefficients.

The method of proof depends upon the observation that, purely formally, partial summation yields

---

Received by the editors January 28, 1944.

[1] Numbers in brackets refer to the references cited at the end of the paper.

$$\sum_1^\infty a_n B_n = -\sum_1^\infty A_n b_n.$$

The justification of this will constitute the principal part of the proof. Three lemmas will be required.

LEMMA 1 [2]. *A necessary and sufficient condition that the series* $\sum_1^\infty a_n \cos nx$ *should be a Fourier series of a function of* L^p, $p>1$, *is that for every* $g(x)$ *belonging to* $L^{p'}$ $(1/p+1/p'=1)$ *with Fourier coefficients* a_n', *the series* $\sum_1^\infty a_n a_n'$ *should be finite* $(C, 1)$.

A series is finite $(C, 1)$ when the arithmetic mean of its partial sums is bounded.

LEMMA 2. *If* $H(x)\sim\sum_1^\infty c_n \cos nx$ *is a function of* L^p, $p>1$, *then*

$$\left|\sum_1^N c_n\right| = O(N^{1/p}).$$

PROOF OF LEMMA 2.

$$\sum_1^N c_n = \frac{1}{\pi}\int_0^\pi H(x)\left[\sum_1^N \cos nx\right]dx.$$

Using Hölder's inequality

$$\begin{aligned}\left|\sum_1^N c_n\right| &\leqq \frac{1}{\pi}\left[\int_0^\pi |H(x)|^p dx\right]^{1/p}\left[\int_0^\pi \left|\sum_1^N \cos nx\right|^{p'} dx\right]^{1/p'} \\ &= O\left[\int_0^\pi \left|\sum_1^N \cos nx\right|^{p'} dx\right]^{1/p'} \\ &= O\left[\int_0^\pi \left|\frac{\sin Nx}{x}\right|^{p'} dx\right]^{1/p'} = O(N^{1/p}).\end{aligned}$$

This finishes the proof of the lemma.

LEMMA 3.[2] *If* $H(x)\sim\sum_1^\infty c_n \cos nx$ *belongs to* $L^{p'}$, $p>1$, *then*

$$\left|\sum_N^\infty \frac{c_n}{n}\right| = O(N^{-1/p}).$$

PROOF OF LEMMA 3. Let $\sum_1^n c_\nu = s_n$, then

$$\sum_n^m \frac{c_\nu}{\nu} = m^{-1}s_m - n^{-1}s_{n-1} + \sum_n^{m-1}\frac{s_\nu}{\nu(\nu+1)}.$$

[2] The above proof of Lemma 3 and a part of the proof of Theorem 2 are due to Professor O. Szász and are simpler and more elegant than the original.

By Lemma 2, $s_m = O(m^{1/p'})$, hence $m^{-1}s_m \to 0$, and

$$\sum_n^\infty \nu^{-1}c_\nu = \frac{-s_{n-1}}{n} + \sum_n^\infty \frac{s_\nu}{\nu(\nu+1)}$$

$$= O(n^{1/p'-1}) + O\left(\sum_n^\infty \nu^{1/p'-2}\right)$$

$$= O(n^{1/p'-1}) = O(n^{-1/p}).$$

PROOF OF THEOREM 2. A_n and B_n, $f(x)$, $g(x)$, $F(x)$, $G(x)$ are defined as in Theorems 1 and 2. We divide the proof into three parts. The first part will consist of showing that $G(x)$ belongs to L^p, $p>1$, whenever $g(x)$ does. In virtue of Lemma 1, it is sufficient to show

$$\sum_1^n B_\nu a_\nu = O(1) \qquad \text{as } n \to \infty,$$

where $g(x) \sim \sum b_n \cos nx$ is a function of L^p, and $f(x) \sim \sum a_n \cos nx$ is a function of $L^{p'}$.

By partial summation

$$\sum_1^n B_\nu a_\nu = B_n \sum_1^n a_\nu - \sum_1^{n-1} A_\nu b_\nu. \tag{1}$$

Using Theorem 1, $F(x) \sim \sum A_n \cos nx$ belongs to $L^{p'}$, and thus, using the generalized Parseval theorem, $\sum_1^\infty A_\nu b_\nu$ is convergent. Furthermore, $\sum_1^n a_\nu = O(n^{1/p'})$, by Lemma 2, and $B_\nu = O(n^{-1/p'})$, by Lemma 3. Hence

$$B_n \sum_1^n a_\nu = O(n^{1/p'})O(n^{-1/p'}) = O(1),$$

and thus $G(x)$ belongs to L^p.

The second part consists of showing that $\sum_1^\infty B_\nu a_\nu$ converges. This is immediate by the generalized Parseval theorem (Hausdorff-Young theorem) since $G(x)$ belongs to L^p, $f(x)$ to $L^{p'}$.

The third part consists of showing that $\sum_1^\infty A_n b_n = -\sum_1^\infty a_n B_n$ and using (1) this is true if, and only if, $B_n \sum_1^n a_\nu = o(1)$.

We have to consider two cases.

Case 1. $f(x)$ belongs to L^p, $1 < p \leqq 2$. Then, by the Hausdorff-Young theorem, $\sum |a_n|^{p'}$ converges, and thus

$$\left| \sum_1^N a_n \right| = o(N^{1/p}).$$

Case 2. $f(x)$ belongs to L^p, $p>2$, and thus $g(x)$ belongs to $L^{p'}$,

$1<p'\leqq 2$. Then, as before, $\sum |b_n|^{p'}$ converges, and this entails

$$\left|\sum_N^\infty \frac{b_n}{n}\right| = o(N^{-1/p}).$$

Thus for all $p>1$,

$$\left|B_n \sum_1^n a_r\right| = \left\{\begin{matrix} o(N^{1/p})O(N^{-1/p}) \\ O(N^{-1/p})o(N^{1/p}) \end{matrix}\right\} = o(1).$$

This finishes the proof of Theorem 2.

It is interesting to see how these results arise naturally in dealing with Fourier integrals. The processes will be purely formal, although they could be justified by similar methods.

$$f(t) = \int_0^\infty F(x) \cos xt dx,$$

$$\int_0^t f(t)dt = \int_0^\infty F(x) \frac{\sin xt}{x} dx = t\int_0^\infty \left(\int_x^\infty \frac{F(u)}{u} du\right) \cos xt\, dx.$$

The last result is obtained by integrating by parts, assuming that the integrated term drops out. Thus

$$\frac{1}{t}\int_0^t f(t)dt = \int_0^\infty \left(\int_x^\infty \frac{F(u)}{u} du\right) \cos xt\, dx.$$

Similarly

$$\int_t^\infty \frac{f(t)}{t} dt = \int_0^\infty F(x)\left[\int_t^\infty \frac{\cos xt}{t}\right] dx$$

$$= \int_0^\infty F(x)\left[\int_{xt}^\infty \frac{\cos u}{u} du\right] dx.$$

Integrating the last expression by parts, and assuming the integrated part vanishes, there results

$$\int_t^\infty \frac{f(t)}{t} dt = \int_0^\infty \left[\frac{1}{x}\int_0^x F(u)du\right] \cos xt\, dx.$$

References

1. G. H. Hardy, *Notes on some points in the integral calculus*, Mesesnger of Mathematics vol. 58 (1928) pp. 50–52.

2. A. Zygmund, *Trigonometrical series*, Warsaw, 1935, 4.63, Lemma, p. 103.

Princeton University

SOME INTEGRAL RELATIONS

H. BATEMAN

Let $u(r)$ have a continuous derivative for positive values of r, then a solution of the integral equation

$$u(r) = \int_0^{\pi/2} \cos(kr \sin a) U(r \cos a) da \tag{1}$$

is given by the formula

$$U(s) = (2/\pi)\left[u(0)ch(ks) + s\int_0^{\pi/2} ch(ks \sin b) u'(s \cos b) db\right]. \tag{2}$$

The mean value theorem for a two-dimensional potential round a circle, due essentially to Parseval, gives the equation

$$1 = (2/\pi)\int_0^{\pi/2} \cos(kr \sin a) ch(kr \cos a) da,$$

and so the relation to be established is

$$u(r) - u(0) = (2/\pi)\int_0^{\pi/2} \cos(kr \sin a) r \cos a \, da \int_0^{\pi/2} ch(kx) u'(y) db,$$

where

$$y = r \cos a \cos b, \quad x = r \cos a \sin b = z \sin c, \quad z = (r^2 - y^2)^{1/2}.$$

When the quantities y and c are taken as new variables of integration the integral on the right becomes

$$(2/\pi)\int_0^r u'(y) dy \int_0^{\pi/2} \cos(kz \cos c) ch(kz \sin c) dc = \int_0^r u'(y) dy$$

and so the formula is established.[1] A proof of the uniqueness of the solution is not easy to obtain by substitution of an integral of type (1) in (2) but may be derived from the relation

Received by the editors March 27, 1944.

[1] The case $k=0$ is reducible to Abel's integral equation and is a special case of an equation $F(x) = \int_0^{\pi/2} f(x \cos a) \sin^{2n+1} a \, da$ found by Poisson in his *Researches on heat conduction*, J. École Polytech. vol. 12 (1823) p. 299. This equation was discussed by J. Liouville, ibid. vol. 15 (1835) pp. 55–60. Professor M. A. Bosoco has kindly recalled my attention to the interesting integral transformation used for this case by O. Schlömilch, Zeitschrift für Mathematik und Physik vol. 2 (1857).

$$u(s^{1/2}) = \frac{1}{2}\int_0^s \cos\, k(s-t)^{1/2} U(t^{1/2})dt/[t(s-t)]^{1/2}$$

and the equivalent relation

$$\int_0^\infty e^{-ss}u(s^{1/2})ds = \frac{1}{2}\int_0^\infty e^{-st}U(t^{1/2})t^{-1/2}dt \int_0^\infty e^{-st}\cos\,(kt^{1/2})t^{-1/2}dt.$$

Thus when $u(r)$ is given for positive values of r the first integral on the right is determined uniquely and so by a simple extension of Lerch's theorem $U(r)$ is unique.

The theory of exponential integrals has been advanced by the recent tabulation[2] of the function $F(x)$ which is defined by the integral

$$F(x) = \int_0^\infty e^{-m^{4/4}}\cos\,(mx)dm \tag{3}$$

$$= \frac{1}{2}\left(\frac{1}{2}\right)^{1/2}\sum_{n=0}^{\infty}(-2x^2)^n\Gamma(n/2+1/4)/(2n)!. \tag{4}$$

The table indicates that $F(x)=0$ for $x=2.435$, 4.778, and 6.550, approximately. This is in accordance with the result of F. Bernstein[3] that $F(x)$ has an infinite number of real zeros.

The asymptotic behavior of $F(x)$ for large complex values of x has been discussed by Burwell[4] who gives, moreover, an asymptotic formula for the large zeros. The asymptotic expansion of $F(x)$ for large positive values of x is rather complicated but it is sufficient to establish the convergence of the integral

$$\int_0^\infty e^{-zt}F(t^{1/4})t^{-1/4}dt = \frac{\pi z^{-3/4}}{2}\exp\,(-z^{-1/2}/4), \qquad R(z)>0, \tag{5}$$

which may be regarded as a generating function of $F(t^{1/4})t^{-1/4}$ in the sense of Abel. In the theory of Haar and others by which the asymptotic behavior of a function for large positive values of t is derived from the behavior of the generating Laplacian integral at points z on the line bounding the region of convergence, the form of the function

$$F(t^{1/4})t^{-1/4}$$

for large positive values of t is characterized by the behavior of the

[2] G. Grimminger, *Velocity and mass distributions resulting from the lateral diffusion of a current in a stratified medium on a rotating earth*, Journal of the Franklin Institute vol. 236 (1943) pp. 413–443, 509–520.

[3] F. Bernstein, Math. Ann. vol. 79 (1918), pp. 265–268. See also G. Pólya, Messenger of Mathematics vol. 52 (1923) pp. 185–188.

[4] W. R. Burwell, Proc. London Math. Soc. (2) vol. 22 (1924) pp. 57–72.

function on the right of (5) at the point $z=0$. The type of singularity is in this case different from that of any tabulated generating function.

The relation (5) was obtained originally by substituting the infinite series for $F(x)$ and integrating term by term.[5] An alternative plan is to consider the function

$$U(x, y) = \int_0^\infty e^{-ym^4} \cos(mx)dm \tag{6}$$

which is a solution of the partial differential equation

$$U_y + U_{xxxx} = 0.$$

The integral

$$u(y, z) = \int_0^\infty e^{-zt} U(t^{1/4}, y)t^{-1/4}dt = 4\int_0^\infty e^{-zx^4} U(x, y)x^2dx \tag{7}$$

is then a solution of the partial differential equation

$$u_y = 360zu + 2160z^2u_z + 1664z^3u_{zz} + 256z^4u_{zzz}$$

and may be identified with the solution

$$u(y, z) = \frac{1}{2}\left(\frac{1}{2}\right)^{1/2} \pi y^{-1/4} z^{-3/4} \exp(-y^{-1/2}z^{-1/2}/8)$$

for U is of the form $(1/x)f(yx^{-4})$ and so $u(y, z)$ is of the form

$$u(y, z) = z^{-1/2}H(yz) = z^{-1/2}H(w),$$

say. If $H(w)=w^{-1/4}S(w^{-1/2})$ the differential equation for $S(s)$ is

$$2s\,dT/ds + T = 0 \quad \text{where} \quad T = 64S''(s) - S.$$

Hence $T=Cs^{-1/2}$ and so $S=Ae^{s/8}+Be^{-s/8}-CP(s)$, where for large values of s the function P has an asymptotic expansion of type

$$P(s) = s^{-1/2}[1 + 1\cdot3(16/s^2) + 1\cdot3\cdot5\cdot7(16/s^2)^2 + \cdots].$$

The solution $S=B\exp(-s/8)$ is of the right type when s is large and w small and the proposed value of the integral seems to be correct.

There are some related integrals involving the functions V_0, V_1, V_2, V_3 defined on p. 114 of my *Partial differential equations of mathemati-*

[5] For a justification of this process reference may be made to G. H. Hardy, Transactions of the Cambridge Philosophical Society vol. 21 (1908) pp. 1–48 and E. C. Titchmarsh, *The theory of functions*, Oxford, 1932, p. 348. The infinite series for $F(x)$ may indeed be expressed as the difference of two series involving positive terms and each of these may be integrated term by term because the resulting series are convergent. I owe the second reference to Morgan Ward.

cal physics. These functions are solutions of the partial differential equation[6]

$$V_y = V_{xxxx}$$

and are represented by the integrals

$$V_0(x, y) = 4\int_0^\infty e^{-4ys^4} \cos sx \, ch \, sx \, ds,$$

$$V_2(x, y) = 4\int_0^\infty e^{-4ys^4} \sin sx \, sh \, sx \, ds,$$

$$V_1(x, y) = 4\int_0^\infty e^{-4ys^4}(\sin sx \, ch \, sx + \cos sx \, ch \, sx)ds,$$

$$V_3(x, y) = 4\int_0^\infty e^{-4ys^4}(\sin sx \, ch \, sx - \cos sx \, sh \, sx)ds.$$

The simplest of the integrals are

$$\int_0^\infty e^{-st}t^{-1/4}V_0(s^{1/4}, y)dt = \pi y^{-1/4}z^{-3/4} \cos (y^{-1/2}z^{-1/2}/8),$$

$$\int_0^\infty e^{-st}t^{-1/4}V_2(s^{1/4}, y)dt = \pi y^{-1/4}z^{-3/4} \sin (y^{-1/2}z^{-1/2}/8).$$

The asymptotic forms of $V_0(x, y)$ and $V_2(x, y)$ for large values of x can be derived from Burwell's results; they can also be regarded as determined by the behavior of the functions on the right in the neighborhood of $z=0$. In the search for functions whose generating functions have prescribed singularities it should be worth while to fully explore the functions provided by solutions of the general equation of Poisson.

Generating functions depending on the error function and the integrals of Fresnel may be derived from the functions now under discussion. Thus

$$\int_0^\infty e^{-st}[F(0) - F(t^{1/4})]t^{-5/4}dt = \frac{\pi z^{-1/4}}{2}\sum_{n=0}^\infty (-z^{-1/4}/4)^n/[(2n+1)(n+1)!],$$

$$\int_0^\infty e^{-st}[V(0, y) - V(t^{1/4}, y)]t^{-5/4}dt = \frac{\pi}{16}(y^5z^2)^{-1/4}\sum_{n=0}^\infty (-8y^{1/2}z^{1/2})^{-n}/[(4n+3)(2n+2)!],$$

[6] This is a special case of the equation $\partial^m V/\partial x^m = \partial^n V/\partial y^n$ discussed by Poisson, J. École Polytech. vol. 12 (1823) pp. 464–474.

$$\int_0^\infty e^{-zt} V_2(t^{1/4}, y) t^{-3/4} dt = \frac{\pi y^{-3/4} z^{-1/4}}{2} \sum_{n=0}^{\infty} (-8y^{1/2}z^{1/2})^{-n} / [(4n+1)(2n+1)!].$$

For the solution of the general equation of Poisson, reference may be made to H. Block, Arkiv för Matematik, Astronomi och Fysik vol. 8 (1913), G. Mikaelsson, Dissertation, Upsala, 1920 and E. Rothe, Math. Zeit. vol. 28 (1928) pp. 48–72.

California Institute of Technology

TWO PROPERTIES OF THE FUNCTION COS x

HERBERT E. ROBBINS

The function $f(x)=A\cos n(x+B)$, where A, B are any real constants and n is an integer, has the properties:

(I) $f(x)$ is real valued for all real x, of period 2π, and continuous.

(II) $f(x)$ is differentiable, and there exist constants a, b such that, for all x,

$$f'(x) = af(x+b).$$

(III) Given any constants a, a', b, b', there exist constants c, d such that, for all x,

$$af(x+a') + bf(x+b') = cf(x+d).$$

The object of this note is to show that, conversely, any function $f(x)$ which has property (I) and either (II) or (III) is necessarily of the form $f(x)=A\cos n(x+B)$. The latter result is used to derive the parallelogram law of addition of forces from certain other basic assumptions.

THEOREM 1. *Let $f(x)$ have properties* (I) *and* (II). *Then there exist constants A, B and an integer n such that $f(x)=A\cos n(x+B)$.*

PROOF. It follows from (II) that $f(x)$ is of class C^∞ and hence, from (I), can be represented by a convergent Fourier series, which, moreover, may be differentiated termwise. Thus for some complex constants k_n,

$$\text{(1)}\qquad \begin{aligned} f(x) &= \sum k_n e^{inx}, \qquad f'(x) = \sum ink_n e^{inx}, \\ f'(x) - af(x+b) &= \sum k_n(in - ae^{inb})e^{inx}. \end{aligned}$$

It follows from (II) that for every integer n,

$$\text{(2)}\qquad k_n(in - ae^{inb}) = 0.$$

If $f(x)\equiv 0$ then the theorem is trivial. Otherwise, there will exist an n for which $k_n\neq 0$. It follows that

$$\text{(3)}\qquad in = ae^{inb}.$$

Taking absolute values we have

$$\text{(4)}\qquad n = \pm a.$$

Thus there can be at most two values of n for which $k_n\neq 0$, and these

Received by the editors March 3, 1944.

values are negatives of one another. Thus for some integer n,

$$f(x) = k_{-n}e^{-inx} + k_n e^{inx}. \tag{5}$$

Since, by (I), $f(x)$ is real valued, it follows that k_{-n} and k_n are complex conjugates, and the proof is complete.

THEOREM 2. *Let $f(x)$ have properties* (I) *and* (III). *Then there exist constants A, B and an integer n such that $f(x) = A \cos n(x+B)$.*

PROOF. Since, by (I), $f(x)$ is continuous and of period 2π, it possesses at least a formal Fourier series,[1]

$$f(x) \sim \sum k_n e^{inx}, \qquad k_n = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}dt. \tag{6}$$

By (III), there exist constants c, d such that the function

$$g(x) = f(x+1) + f(x) - cf(x+d), \tag{7}$$

continuous and of period 2π, is identically zero. (This is the only consequence of (III) that we shall use.) Hence

$$0 \sim \sum k_n(e^{in} + 1 - ce^{ind})e^{inx}. \tag{8}$$

It follows that for every n,

$$k_n(e^{in} + 1 - ce^{ind}) = 0. \tag{9}$$

If $f(x) \equiv 0$ then the theorem is trivial. Otherwise, there will exist an n for which $k_n \neq 0$. For any such n it follows from (9) that

$$e^{in} + 1 = ce^{ind}. \tag{10}$$

Taking absolute values and squaring, it follows that

$$\cos n = (c^2 - 2)/2. \tag{11}$$

Hence if m and n are any two integers for which $k_m \cdot k_n \neq 0$, it follows from (11) that $\cos m = \cos n$. Hence for some integer r,

$$m = \pm n + 2\pi r. \tag{12}$$

Since π is irrational, it follows that $r = 0$ and $m = \pm n$. Thus the formal Fourier series for $f(x)$ consists of only two terms,

$$f(x) \sim k_{-n}e^{-inx} + k_n e^{inx}. \tag{13}$$

But in this case, since the functions e^{inx} are complete with respect

[1] The proof given here follows a suggestion of Paul R. Halmos. The author's original proof required the unnecessary assumption that $f(x)$ be of class C^1.

to continuous functions, the relation $\sim$ can be replaced by an identity,

$$f(x) = k_{-n}e^{-inx} + k_n e^{inx}. \tag{14}$$

Since $f(x)$ is real valued, k_{-n} and k_n must be complex conjugates, and the theorem is proved.

We shall now apply Theorem 2 to derive the law of addition of forces.[2] For simplicity, let us consider only forces acting at a fixed point in a fixed plane in which the angular coordinate x is defined. With such a force we identify the real valued function $F(x)$ which specifies the scalar component of the force in the direction x; thus a force is represented by a real valued function of period 2π. By the sum of two forces $F_1(x)$ and $F_2(x)$ we mean the function $F_1(x)+F_2(x)$. Our assumptions are the following.

(i) *All forces are geometrically similar.* By this we mean that there exists a fixed function $f(x)$ of period 2π such that any force $F(x)$ can be written in the form

$$F(x) = A_F \cdot f(x + \alpha_F), \tag{15}$$

where A_F and α_F are constants determined by $F(x)$. We need not assume that all values of the constants A_F and α_F can occur in (15), but we shall assume that there exist at least the forces $F_1(x)=f(x)$ and $F_2(x)=f(x+1)$.

(ii) *The sum of two forces is a force.* Together with (i), this implies that the function $f(x)$ has the property that for certain constants c, d and for every x,

$$f(x + 1) + f(x) = cf(x + d). \tag{16}$$

(iii) *The function $f(x)$ is continuous, non-constant, and vanishes for at most two values in the interval $0 \leqq x < 2\pi$.*

The proof of Theorem 2 shows that the function $f(x)$, continuous, real valued, of period 2π, and satisfying (16), must be of the form

$$f(x) = A \cos n(x + B), \tag{17}$$

where n is an integer. The hypotheses of (iii) ensure that n can be chosen as 1. The parallelogram law of addition of forces is an immediate consequence.

UNITED STATES NAVAL ACADEMY, POST GRADUATE SCHOOL

[2] See G. Darboux, Bull. Sci. Math. vol. 9 (1875) pp. 281–288; also G. D. Birkhoff, Rice Institute Pamphlet vol. 28, no. 1 (1941) pp. 46–50.

ON SOME SPECIAL DIOPHANTINE EQUATIONS

E. ROSENTHALL

1. **Introduction.** The following lemma is fundamental in the algorithm of reciprocal arrays for the solution of multiplicative diophantine equations in certain arithmetics as developed by E. T. Bell.[1]

LEMMA 1. *All sets of integers satisfying the diophantine equation*

$$xy = zw$$

are given by

$$x = ab, \quad y = cd, \quad z = ad, \quad w = cb$$

and it suffices to take b and d coprime.

Rational arithmetic and the theory of ideals in an algebraic number field provide instances of these arithmetics, and in all cases the fundamental theorem of unique decomposition into prime factors is required.

By use of this algorithm Bell has obtained the complete solution of a large class of diophantine equations, and by means of an application of Lemma 1 (and results derived from it) to equations reducible in particular algebraic number fields the present writer has obtained the complete solution of some interesting diophantine equations.

In this paper Lemma 1 is generalized to an arbitrary algebraic field, and the method of proof is then applied to a multiplicative equation from which we immediately obtain a formula exhibiting all the rational integers satisfying $x^2+ay^2=z^{2n+1}$. The procedure is to replace the algebraic indeterminates in the given multiplicative equation by the principal ideals they generate; then solving this equation by the method of arrays we obtain the solution in terms of ideals. In this solution, by a use of properties of equivalent ideals all the ideals are replaced by suitable principal ideals and the complete solution of the given equation is deduced.

2. **Notations.** We shall adhere to the following notations: small letters $a, b, \cdots$ are reserved for the rational integers, the capital letters $A, B, \cdots$ (except E) for integers of the algebraic number field $\mathfrak{F}$; the letter ϵ will be reserved for the units of this field, and all other Greek letters (with or without subscripts) will denote ideals of $\mathfrak{F}$.

Received by the editors March 24, 1944.

[1] E. T. Bell, *Reciprocal arrays and diophantine analysis*, Amer. J. Math. vol. 35 (1933) pp. 50–66.

Parentheses enclosing a letter denote the corresponding principal ideal; thus (X), (e), $\cdots$. The conjugates of the ideal ξ are represented by ξ', ξ'', $\cdots$, $\xi^{(n-1)}$. Two ideals, α and β, are said to be equivalent if an ideal γ exists such that the products $\alpha\gamma$, $\beta\gamma$ are both principal ideals; the equivalence of α and β is expressed by $\alpha\sim\beta$.

3. **Generalization of Lemma 1.** We shall now prove the following result.

THEOREM 1. *All integers X, Y, Z, W satisfying*

$$XY = ZW \tag{3.1}$$

are given by

$$X = ST/e, \qquad Y = UV/e, \qquad Z = SV/e, \qquad W = UT/e,$$

where S, T, U, V are arbitrary and e takes only the finite set of rational integral values each equal to the norm of a representative ideal from each class.

PROOF. By Lemma 1 all solutions of

$$(X)(Y) = (Z)(W) \tag{3.2}$$

are obtainable from

$$(X) = \alpha\beta, \qquad (Y) = \gamma\delta, \qquad (Z) = \alpha\delta, \qquad (W) = \gamma\beta,$$

and the ideals of the right-hand members must be restricted to the values which make the left-hand members principal ideals. Since the products $\alpha\beta$, $\gamma\beta$ are to be principal ideals then $\alpha\sim\gamma$. Let ξ be a representative ideal of each class in which α is a member. Then $\alpha\sim\xi$ and it follows that

$$\alpha\xi'\xi'' \cdots \xi^{(n-1)} \sim \xi\xi' \cdots \xi^{(n-1)} = (e).$$

Hence $\alpha\xi'\xi'' \cdots \xi^{(n-1)}$, and consequently $\gamma\xi'\xi'' \cdots \xi^{(n-1)}$, is equivalent to a principal ideal and must therefore itself be a principal ideal (A). Therefore,

$$(eX) = \alpha\beta\xi\xi' \cdots \xi^{(n-1)} = (A)\xi\beta$$

whence $\xi\beta$ is a principal ideal. Put $\xi\beta = (B)$ and we have $(eX) = (AB)$. Also

$$(eY) = \xi\xi' \cdots \xi^{(n-1)}\gamma\delta = (C)\xi\delta = (C)(D)$$

since $\xi\delta$ must be a principal ideal (D).

Thus all solutions of (3.2) are obtainable from

$$(eX) = (AB), \qquad (eY) = (CD), \qquad (eZ) = (AD), \qquad (eW) = (CB).$$

This implies that the set of all integers satisfying (3.1) is given by

$$eX = \epsilon_1 AB, \qquad eY = \epsilon_2 CD, \qquad eZ = \epsilon_3 AD, \qquad eW = \epsilon_4 CB,$$

where $\epsilon_1\epsilon_2 = \epsilon_3\epsilon_4$. Make the reversible substitution $A = \epsilon_1^{-1}S$, $B = T$, $D = \epsilon_1\epsilon_3^{-1}V$, $C = \epsilon_4^{-1}U$ and we have the required result.

4. **The equation $X\overline{X} = z^{2n+1}$.** Hereafter the field $\mathfrak{F}$ is an arbitrary quadratic number field and the conjugate of an integer X and an ideal α are denoted by $\overline{X}$ and $\bar{\alpha}$ respectively.

The following theorem will be verified by induction.

THEOREM 2. *All solutions of*

$$\alpha\bar{\alpha} = (z^{2n+1})$$

are given by

$$\alpha = \phi_1^{2n+1}\phi_2^{2n}\bar{\phi}_2\phi_3^{2n-1}\bar{\phi}_3^{2} \cdots \phi_{n+1}^{n+1}\bar{\phi}_{n+1}^{n}, \qquad (z) = \phi_1\bar{\phi}_1\phi_2\bar{\phi}_2 \cdots \phi_{n+1}\bar{\phi}_{n+1}.$$

From this we can deduce the following result.[2]

THEOREM 3. *All solutions of*

$$X\overline{X} = z^{2n+1} \tag{4.1}$$

are given by

$$\begin{aligned} E^{2n+1}X &= \pm H_1^{2n+1}H_2^{2n}\overline{H}_2 \cdots H_{n+1}^{n+1}\overline{H}_{n+1}^{n} \\ E^{2}z &= H_1\overline{H}_1H_2\overline{H}_2 \cdots H_{n+1}\overline{H}_{n+1}, \end{aligned} \tag{4.2}$$

where $E = e_1^{n+1}e_2^{n} \cdots e_n^{2}$; $e_1, e_2, \cdots, e_n$ each being equal to the norm of a representative ideal from each class.

If we put $X = x + (-a)^{1/2}y$, then (4.1) becomes $x^2 + ay^2 = z^{2n+1}$ and equating rational and irrational parts we obtain from (4.2) an explicit representation for all rational integers satisfying this equation. Although in general the solution appears in rational form, yet for each value of a the indeterminates x, y, z can be expressed by a finite number of polynomials in integral parameters. For, for each value of a the parameter E has only a finite set of integral values and the requirement that the right-hand members of (4.2) be divisible by

[2] For an account of the investigations on equation $x^2 + ay^2 = z^n$ see L. E. Dickson, *History of the theory of numbers*, vol. 2, pp. 534–543; Th. Skolem, *Diophantische Gleichungen*, Berlin, 1938, pp. 64–68; J. V. Uspensky and M. A. Heaslet, *Elementary number theory*, 1939, pp. 389–396.

E^{2n+1}, E^2 respectively may be expressed by congruential conditions upon the coordinates of the integers H_i.

EXAMPLE. From Theorem 3 it follows that all the rational integers satisfying $x^2+47y^2=z^3$ are given by

$$E^3(x+(-47)^{1/2}y) = \pm H_1^3H_2^2\overline{H}_2, \qquad E^3z = H_1\overline{H}_1H_2\overline{H}_2,$$

where $E=e_1^2=1$, 2^2, 3^2 and H_1, H_2 are integers of $Ra(-47)^{1/2}$. It suffices to take H_2 primitive.

In order to select all integers from the above rational forms of x, y, z, it is necessary and sufficient to impose the following congruential conditions upon the coordinates of $H_1=r+sW$, $H_2=m+nW$ where $W=(1+(-47)^{1/2})/2$:

(i) $E=1$. If n even then r, s even; otherwise no restrictions on r, s.

(ii) $E=2^2$. m odd, n even, $s\equiv 2r\equiv 0 \pmod 8$; $m-4n\equiv 2 \pmod 4$, $r\equiv s\equiv 0 \pmod 4$; $m-4n\equiv 4 \pmod 8$, $r\equiv s\equiv 0, 2 \pmod 4$; $m-4n\equiv 8 \pmod{16}$, $r\equiv s\equiv 0, 2 \pmod 4$; $m-4n\equiv 0, \pm 16 \pmod{64}$, $r\equiv s \pmod 2$; $m-4n\equiv 32 \pmod{64}$ requires no restriction on r, s.

m odd, $m+5n\equiv 2 \pmod 4$, $r\equiv s\equiv 0 \pmod 4$; $m+5n\equiv 4 \pmod 8$, $r\equiv s\equiv 0 \pmod 4$; $m+5n\equiv 8 \pmod{16}$, $r\equiv 2s \pmod 4$; $m+5n\equiv 0, \pm 16 \pmod{64}$, r even; $m+5n\equiv 32 \pmod{64}$ requires no restriction on r and s.

(iii) $E=3^2$. $m(m+n)\not\equiv 0 \pmod 3$, $r\equiv s\equiv 0 \pmod 9$; $m-6n\equiv \pm 3 \pmod 9$, $r\equiv s\equiv 0 \pmod 9$; $m-6n\equiv \pm 9 \pmod{27}$, $r+s\equiv 0 \pmod 9$; $m-6n\equiv 0 \pmod{27}$, but $m-114n\not\equiv 0 \pmod{3^6}$, $r+s\equiv 0 \pmod 3$; $m-114n\equiv 0 \pmod{3^6}$ requires no restriction on r and s.

$m+7n\equiv \pm 3 \pmod 9$, $r\equiv s\equiv 0 \pmod 9$; $m+7n\equiv \pm 9 \pmod{27}$, $r\equiv 3s\equiv 0 \pmod 9$; $m+7n\equiv 0 \pmod{27}$ but $m+115n\not\equiv 0 \pmod{3^6}$, $r\equiv 0 \pmod 3$; $m+115n\equiv 0 \pmod{3^6}$ requires no restriction on r and s.

If in addition to the above n is even then r and s must both be even.

The above conditions were obtained by considering in turn the cases according as $H_2\overline{H}_2$ is divisible by $e_1, e_1^2, \cdots, e_1^6$.

For the proof of Theorem 2 we use the following two lemmas.

LEMMA 2. *All solutions of* $\alpha\bar{\alpha}=\gamma\beta\bar{\beta}$ *are given by* $\alpha=\theta_1\theta_2\theta_3$, $\beta=\theta_1\bar{\theta}_2$, $\gamma=\theta_3\bar{\theta}_3$.

LEMMA 3. *All solutions of* $\alpha\bar{\alpha}=\beta\gamma$ *are given by* $\alpha=\theta_1\theta_2\theta_3\theta_4$, $\beta=\theta_1\bar{\theta}_2\theta_3\bar{\theta}_3$, $\gamma=\bar{\theta}_1\theta_2\theta_4\bar{\theta}_4$.

The proof for Lemma 2 is exactly as given in a previous paper[3] where the indeterminates were integers of a unique factorization

[3] E. Rosenthall, *On some cubic diophantine equations*, Amer. J. Math. vol. 65 (1943) pp. 664–665.

quadratic field; Lemma 3 follows immediately by a repeated application of Lemma 1.

5. **Proof of Theorem 2.** We now show independently that Theorem 2 holds for $n=1$ and then complete the proof by mathematical induction. From Lemma 2 it follows that all α, (z) satisfying $\alpha\bar{\alpha}=(z^2)$ are obtainable from

$$\alpha = \theta_1\theta_2\theta_3, \qquad (z) = \theta_1\bar{\theta}_2 = \theta_2\bar{\theta}_3,$$

and applying Lemma 3 it follows that

$$\theta_3 = \lambda_1\lambda_2\lambda_3\lambda_4, \qquad \theta_1 = \lambda_1\bar{\lambda}_2\lambda_3\bar{\lambda}_3, \qquad \bar{\theta}_2 = \bar{\lambda}_1\lambda_2\lambda_4\bar{\lambda}_4.$$

Then

$$\alpha = \phi_1^3\phi_2^2\bar{\phi}_2, \qquad (z) = \phi_1\bar{\phi}_1\phi_2\bar{\phi}_2,$$

where we have put $\lambda_1=\phi_1$, $\bar{\lambda}_2\lambda_3\lambda_4=\phi_2$ since λ_2, λ_3, λ_4 always appear in this product form.

Now consider the induction from $n=s$ to $n=s+1$. Apply Lemma 2 to $\alpha\bar{\alpha}=(z^{2s+2})=(z^{2s+1})(z)(z)$. Then

$$\alpha = \theta_1\theta_2\theta_3, \qquad (z) = \theta_1\bar{\theta}_2, \qquad (z^{2s+1}) = \theta_2\bar{\theta}_3.$$

By hypothesis, Theorem 2 holds for the last equation. Hence

$$\theta_3 = \mu_1^{2s+1}\mu_2^{2s}\bar{\mu}_2\cdots\mu_{s+1}^{s+1}\bar{\mu}_{s+1}^{s}, \qquad (z) = \mu_1\bar{\mu}_1\mu_2\bar{\mu}_2\cdots\mu_{s+1}\bar{\mu}_{s+1};$$

and equating the two parametric representations for (z) we obtain the following equations for all the parameters concerned,

$$\{\mu_1\mu_2\cdots\mu_{s+1}\}\{\bar{\mu}_1\bar{\mu}_2\cdots\bar{\mu}_{s+1}\} = \theta_1\bar{\theta}_2.$$

Applying Lemma 3,

$$\mu_1\mu_2\cdots\mu_{s+1} = \lambda_1\lambda_2\lambda_3\lambda_4, \qquad \theta_1 = \lambda_1\bar{\lambda}_2\lambda_3\bar{\lambda}_3, \qquad \bar{\theta}_2 = \bar{\lambda}_1\lambda_2\lambda_4\bar{\lambda}_4.$$

Substituting these values of the parameters in the preceding formulas for α, (z) gives

$$(5.1) \qquad \alpha = \lambda_1^{2s}\bar{\lambda}_2^{2}\psi_1\bar{\psi}_1\mu_1^{2s+1}\mu_2^{2s}\bar{\mu}_2\cdots\mu_{s+1}^{s+1}\bar{\mu}_{s+1}^{s}, \qquad (z) = \lambda_1\bar{\lambda}_1\lambda_2\bar{\lambda}_2\psi_1\bar{\psi}_1,$$

where the parameters must satisfy the multiplicative equation

$$(5.2) \qquad \mu_1\mu_2\cdots\mu_{s+1} = \lambda_1\lambda_2\psi_1,$$

where we have put $\lambda_3\lambda_4=\psi_1$.

Applying the method of arrays[4] to (5.2) we obtain

[4] See E. T. Bell, *loc. cit.*, p. 62.

$$\mu_1 = \xi_1\eta_1\zeta_1, \qquad \lambda_1 = \xi_1\xi_2\cdots\xi_{s+1},$$
$$\mu_2 = \xi_2\eta_2\zeta_2, \qquad \lambda_2 = \eta_1\eta_2\cdots\eta_{s+1},$$
$$\cdots\cdots\cdots, \qquad \psi_1 = \zeta_1\zeta_2\cdots\zeta_{s+1}.$$
$$\cdots\cdots\cdots,$$
$$\mu_{s+1} = \xi_{s+1}\eta_{s+1}\zeta_{s+1}.$$

Finally, substituting in (5.1) gives

$$\alpha = \phi_1^{2s+3}\phi_2^{2s+2}\bar{\phi}_2\cdots\phi_{s+1}^{s+3}\bar{\phi}_{s+1}^{s}\phi_{s+2}^{s+2}\bar{\phi}_{s+2}^{s+1}, \qquad (z) = \phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\cdots\phi_{s+2}\bar{\phi}_{s+2},$$

where we have put $\xi_1=\phi_1$, $\xi_2\zeta_1=\phi_2$, $\zeta_{s+1}\eta_s\bar{\eta}_{s+1}=\phi_{s+2}$, and $\xi_i\zeta_{i-1}\eta_{i-2}=\phi_i$ for $i=3, 4, \cdots, s+1$, since the parameters always occur in these product forms. This completes the induction.

6. **Proof of Theorem 3.** All solutions of

$$(X)(\bar{X}) = (z^{2n+1}) \tag{6.1}$$

are obtainable from

$$(X) = \phi_1^{2n+1}\phi_2^{2n}\bar{\phi}_2\cdots\phi_{n+1}^{n+1}\bar{\phi}_{n+1}^{n}, \qquad (z) = \phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\cdots\phi_{n+1}\bar{\phi}_{n+1}, \tag{6.2}$$

and the ideals of the right-hand members must be restricted to those values which make the left-hand members principal ideals. From $(6.2)_1$ it follows that $\phi_1^{2n+1}\phi_2^{2n-1}\cdots\phi_{n+1}\sim(1)$. Let ξ_i $(i=1, 2, \cdots, n)$ be a representative ideal from each of the classes in which ϕ_i is a member. Then we can put $\phi_i\xi_i=(H_i)$ and $\xi_1^{2n+1}\xi_2^{2n-1}\cdots\xi_n^3\phi_{n+1}=(H_{n+1})$. Multiplying both sides of $(6.2)_1$ by $\left[\left\{\xi_1\bar{\xi}_1\right\}^{n+1}\left\{\xi_2\bar{\xi}_2\right\}^{n}\cdots\left\{\xi_n\bar{\xi}_n\right\}^{2}\right]^{2n+1}$ and $(6.2)_2$ by $\left[\left\{\xi_1\bar{\xi}_1\right\}^{n+1}\left\{\xi_2\bar{\xi}_2\right\}^{n}\cdots\left\{\xi_n\bar{\xi}_n\right\}^{2}\right]^{2}$ and also putting $\xi_i\bar{\xi}_i=(e_i)$ we obtain

$$(E^{2n+1}X) = (H_1^{2n+1}H_2^{2n}\bar{H}_2\cdots H_{n+1}^{n+1}\bar{H}_{n+1}^{n}),$$
$$(E^{2}z) = (H_1\bar{H}_1H_2\bar{H}_2\cdots H_{n+1}\bar{H}_{n+1})$$

for all solutions of (6.1). This implies that all solutions of (4.1) are given by

$$E^{2n+1}X = \epsilon_1H_1^{2n+1}H_2^{2n}\bar{H}_2\cdots H_{n+1}^{n+1}\bar{H}_{n+1}^{n}, \qquad E^{2}z = \epsilon_2H_1\bar{H}_1\cdots H_{n+1}\bar{H}_{n+1}.$$

Make the reversible substitution $H_{n+1}=\bar{\epsilon}_1H_{n+1}$, and then observe that $\epsilon_2\epsilon_1\bar{\epsilon}_1$ must be unity. Hence Theorem 3 is proved.

McGill University

A NOTE ON CHAIN CONDITIONS IN NILPOTENT RINGS AND GROUPS

S. A. JENNINGS

Maximal and minimal conditions for ideals in associative rings have often been considered, but little seems to be known of these conditions in non-associative rings, or of chain conditions on the non-normal subgroups of a group. Moreover, it is usual to assume the condition for one-sided ideals in noncommutative rings, and the weaker condition for two-sided ideals rarely appears. In this note we first consider a class of groups which are "nilpotent" with respect to a set of operators Ω. These groups include ordinary nilpotent groups, and associative and non-associative nilpotent rings and algebras as special cases. Our main theorem is to the effect that, for an Ω-nilpotent group, the maximal or minimal condition for Ω-subgroups implies the corresponding condition for all subgroups. As immediate consequences of this theorem it follows that, for nilpotent rings and algebras, the maximal or minimal condition for ideals implies the corresponding condition for modules, while for nilpotent groups, the maximal or minimal condition for normal subgroups implies the corresponding condition for all subgroups.

1. **Definition of Ω-nilpotency.** Let R be a group, and let Ω be a set of (left) operators of R which includes the inner automorphisms of R. R may also admit a second set of operators Φ, in which case it will be understood in what follows that all subgroups are supposed to admit Φ. The statement that a certain subgroup is an *Ω-subgroup* means that this subgroup admits Ω as well as Φ.

If A is any Ω-subgroup of R, A is normal, and we may suppose that R/A admits the same sets of operators, Ω and Φ, as R.

We take the following as our

DEFINITION OF AN Ω-NILPOTENT GROUP.[1] *R is Ω-nilpotent if there exists a strictly decreasing chain of Ω-subgroups*

Presented to the Royal Society of Canada, May 25, 1943, and to the American Mathematical Society, April 29, 1944; received by the editors January 21, 1944, and, in revised form, March 18, 1944.

[1] Originally I discussed the chain theorems for groups for and rings separately, and while it was clear that the results were analogous, I did not succeed in unifying the proofs for the two cases. The referee, Dr. Irving Kaplansky, pointed out, however, that this could be accomplished by the use of the general notion of Ω-nilpotency which we give here, and which was suggested by him. I would like to express my thanks to the referee for his contribution to this paper.

(1) $$R = A_1 \supset A_2 \supset A_3 \supset \cdots \supset A_m \supset A_{m+1} = 1,$$

terminating with the identity, and such that if $\lambda \in \Omega$ *and* $a_i \in A_i$ *then*

(2) $$a_i^{-r}(\lambda a_i) \in A_{i+1}, \qquad i = 1, 2, \cdots, m,$$

for some r*, where in general* r *may be any positive or negative integer, or zero, but if* λ *happens to be an inner automorphism then* r *must equal one.*

The minimum length of all chains satisfying (1) and (2) above will be called the *exponent* of R and any chain of Ω-subgroups satisfying these conditions and having minimum length will be called a *minimum chain.*

The following results are needed for our later induction.

LEMMA A. *If* R *is an* Ω*-nilpotent group of exponent* m*, and*

$$R = A_1 \supset A_2 \supset \cdots \supset A_m \supset A_{m+1} = 1$$

is a minimum chain, then R/A_m *is an* Ω*-nilpotent group of exponent* $m-1$.

LEMMA B. *The subgroup* A_m *of any chain satisfying* (1) *and* (2) *above is contained in the centre of* R*. Moreover, every subgroup of* A_m *is an* Ω*-subgroup of* R.

LEMMA C. *If* A *is any* Ω*-subgroup of* R*, then the maximal or minimal condition for* Ω*-subgroups of* R *implies the corresponding condition for* Ω*-subgroups in* R/A.

Lemmas A and C may be readily verified. To prove Lemma B, we note that if $a_m \in A_m$, $\lambda \in \Omega$, then $a_m^{-r}(\lambda a_m) = 1$, since $A_{m+1} = 1$, and hence $\lambda a_m = a_m^r$, which implies that every subgroup of A_m admits Ω. Again, if $x \in R$, $x^{-1}a_m x = a_m$, since for inner automorphisms $r = 1$, and hence very element a_m of A_m is in the centre of R.

2. **The chain theorem.** We are now in a position to prove the following

THEOREM. *Let* R *be an* Ω*-nilpotent group. The maximal or minimal condition for* Ω*-subgroups in* R *implies the maximal or minimal condition respectively for all subgroups in* R.

We prove that if the maximal conditions for Ω-subgroups holds in R, the maximal condition for all subgroups also holds; the corresponding result for the minimal condition may be proved in a similar way.

Suppose that R has exponent m, and satisfies the maximal condi-

tion for Ω-subgroups. If $m=1$, our result is true by Lemma B. We proceed by induction and assume that the result holds for groups with exponent at most $m-1$, and hence, by Lemma C and Lemma A, for $\overline{R}=R/A_m$. Let

$$M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$$

be an ascending chain of subgroups of R, and let $\overline{M}_i=(M_i \cup A_m)/A_m$ be the image in $\overline{R}$ of M_i. We have clearly

$$\overline{M}_1 \subseteq \overline{M}_2 \subseteq \overline{M}_3 \subseteq \cdots,$$

an ascending chain of subgroups in $\overline{R}$. Now the maximal condition for subgroups holds in $\overline{R}$, since the condition for Ω-subgroups holds in $\overline{R}$ by Lemma C, and hence there exists an integer j such that

$$\overline{M}_i = \overline{M}_{i+1} \qquad \text{for } i \geqq j.$$

This implies that

$$M_i \cup A_m = M_{i+1} \cup A_m \qquad \text{for } i \geqq j.$$

Consider $N_i=M_i \cap A_m$; clearly N_i is a subgroup of A_m, and hence by Lemma B is an Ω-subgroup of R, and we have

$$N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots,$$

an ascending chain of Ω-subgroups of R. Hence there exists an integer k such that

$$N_i = N_{i+1} \qquad \text{for } i \geqq k.$$

Under our hypothesis, therefore, we see that, provided $i \geqq \max(j, k)$,

$$\begin{aligned} &M_i \subseteq M_{i+1}, \\ (3) \qquad &M_i \cup A_m = M_{i+1} \cup A_m, \\ &M_i \cap A_m = M_{i+1} \cap A_m. \end{aligned}$$

Now A_m belongs to the centre of R, and hence M_i is normal in $M_i \cup A_m$, and therefore normal in $M_{i+1} \cup A_m$: also M_{i+1} is normal in $M_{i+1} \cup A_m$. Therefore the five subgroups M_i, M_{i+1}, $M_i \cup A_m = M_{i+1} \cup A_m$, $M_i \cap A_m = M_{i+1} \cap A_m$, A_m are all normal subgroups of $M_{i+1} \cup A_m$, and hence form a modular lattice, and hence the conditions (3) above imply $M_i = M_{i+1}$, which proves our theorem.

3. **The chain theorem of nilpotent groups.** Consider now the case where R is a nilpotent group in the usual sense.[2] It is easily seen that, if Ω is the set of inner automorphisms, R is Ω-nilpotent. For let

[2] For definitions and properties of nilpotent groups see, for example, H. Zassenhaus, *Gruppentheorie*, vol. 1, Leipzig, 1937, pp. 118–119.

$$R = A_1 \supset A_2 \supset \cdots \supset A_m \supset A_{m+1} = 1$$

be any strictly decreasing central series of R (for example, the lower central series); we have always, if $a_i \in A_i$, $x \in R$,

$$(a_i^{-1} x^{-1} a_i x) \in A_{i+1},$$

which is precisely condition (2) above applied to an inner automorphism, with $r=1$. Since the Ω-subgroups are precisely the normal subgroups of R, we have:

"*In a nilpotent group the maximal or minimal condition for normal subgroups implies the corresponding condition for all subgroups.*"

4. **The chain theorem for nilpotent rings.** The definition of nilpotency is not well established for general non-associative rings. We give here a definition which applies to any ring, associative or not, and which is equivalent to the familiar one in the case of associative and Lie rings. This particular definition seems natural, and the definition of an Ω-nilpotent group as given earlier is a further generalization.

We shall say that R is a generalized ring, or briefly a *ring*, which admits Φ as a set of (left) operators if the following conditions hold:

(a) R is an abelian group which admits Φ.

(b) If $a, x \in R$, then $ax \in R$, that is, R is closed under multiplication.

(c) If $a, b, x \in R$, then $x(a+b) = xa + xb$, $(a+b)x = ax + bx$, that is, the multiplications are left and right operators of R.

It is clear that associative and Lie rings and algebras, and hypercomplex systems both associative and non-associative, are rings in the above sense. We may also suppose that Φ is not empty, since the rational integers may always be adjoined to any domain of operators. By a *module* of R we mean an additive subgroup of R which admits Φ, that is, a module is a Φ-subgroup. By a *left, right or two-sided ideal* of R we mean a module which admits the elements of R as left, right or two-sided operators respectively. Unless otherwise stated, we shall understand by ideal a two-sided ideal.

If A is any ideal of R, the residue classes modulo A form a generalized ring R/A which is homomorphic to R, and we may suppose that R and R/A have the same set of operators.

Because our rings are not necessarily associative, we take the following as our

DEFINITION OF A NILPOTENT RING. *A generalized ring R will be said to be nilpotent if it has a strictly decreasing chain of ideals*

$$R = A_1 \supset A_2 \supset A_3 \supset \cdots \supset A_m \supset A_{m+1} = 0,$$

terminating with zero, such that if $x \in R$ *and* $a_i \in A_i$ *then*

$$xa_i \in A_{i+1}, \qquad a_i x \in A_{i+1}.$$

If R is an associative or Lie ring or algebra, the above definition of nilpotency is obviously equivalent to the usual one.

It is clear that a generalized ring which is nilpotent as defined above is an Ω-nilpotent abelian group, where Ω consists of the right and left multiplications of R, and the integer r of condition (2) above always equals zero. The Ω-subgroups are precisely the ideals of R, and our earlier theorem implies that

"*In any generalized nilpotent ring (and in particular, in any nilpotent associative or Lie ring or algebra) the maximal or minimal condition for ideals implies the corresponding condition for modules.*"

THE UNIVERSITY OF BRITISH COLUMBIA

SUBDIRECT UNIONS IN UNIVERSAL ALGEBRA

GARRETT BIRKHOFF

1. **Preliminary definitions.** By an *algebra*, we shall mean below any collection A of elements, combined by any set of single-valued operations f_ω,

$$y = f_\omega(x_1, \cdots, x_{n(\omega)}). \tag{1}$$

The number of distinct operations (that is, the range of the variable ω) may be infinite, but for our main result (Theorem 2), we shall require every $n(\omega)$ to be finite—that is, it will concern *algebras with finitary operations*.

The concepts of *subalgebra*, *congruence relation* on an algebra, *homomorphism* of one algebra A onto (or into) another algebra with the same operations, and of the *direct union* $A_1 \times \cdots \times A_r$ of any finite or infinite class of algebras with the same operations have been developed elsewhere.[1] More or less trivial arguments establish a many-one correspondence between the congruence relations θ_i on an algebra A and the homomorphic images $H_i = \theta_i(A)$ of the algebra (isomorphic images being identified); moreover the congruence relations on A form a lattice (the *structure lattice* of A). In this lattice, the equality relation will be denoted 0; all other congruence relations will be called *proper*.

More or less trivial arguments also show (cf. *Lattice theory*, Theorem 3.20) that the isomorphic representations of any algebra A as a *subdirect union*, or subalgebra $S \leqq H_1 \times \cdots \times H_r$ of a direct union of algebras H_i, correspond essentially one-one to the sets of congruence relations θ_i on A such that $\Delta\theta_i = 0$. In fact, given such a set of θ_i, the correspondence

$$\theta: \quad a \to [\theta_1(a), \cdots, \theta_r(a)] = [h_1, \cdots, h_r] \tag{2}$$

exhibits the desired isomorphism of A with a subalgebra of $H_1 \times \cdots \times H_r$, where $H_i = \theta_i(A)$. Incidentally, the number of S_i can

Presented to the Society, April 29, 1944, under the title *Subdirect products in universal algebra*; received by the editors March 10, 1944, and, in revised form, June 5, 1944.

[1] *On the structure of abstract algebras*, Proc. Cambridge Philos. Soc. vol. 31 (1935) pp. 433–454, and in the foreword to the author's *Lattice theory*. The idea of an abstract congruence relation is also developed in chap. VI, §14, of S. MacLane's and the author's *Survey of modern algebra*. Interesting remarks in this connection may be found in J. C. C. McKinsey's and A. Tarski's *The algebra of topology*, Ann. of Math. vol. 45 (1944) esp. pp. 190–191.

be infinite. What is more important, the operations of A need not even be finitary.[2]

Equally trivial arguments extend a well known theorem of Emmy Noether on commutative rings[3] to abstract algebras in general. In order to state this extension, we first define an algebra A to be *subdirectly irreducible* if in any finite or infinite representation (2), some θ_i is itself an isomorphism. This means that the meet θ^* of all proper congruence relations on A is itself a proper congruence relation. In lattice-theoretic language, it means that the structure lattice $L(A)$ of A contains a *point* $\theta^*>0$ such that $\theta>0$ implies $\theta\geqq\theta^*$. Hence if A is subdirectly irreducible, $\theta\cap\theta'=0$ in $L(A)$ implies $\theta=0$ or $\theta'=0$; such an A we shall call *weakly irreducible*. If $L(A)$ satisfies the *descending* chain condition, and A is weakly irreducible, then it is evidently also subdirectly irreducible in the strong sense.

If $L(A)$ satisfies the *ascending* chain condition, then it is evident by induction[3] that there exists a representation of 0 as the meet $0=\theta_1\cap\cdots\cap\theta_r$ of a finite number of irreducible elements. This yields the following easy generalization of Emmy Noether's theorem on commutative rings.

THEOREM 1. *Any algebra A whose structure lattice satisfies the ascending chain condition is isomorphic with a subdirect union of a finite number of weakly irreducible algebras.*

For this result, we still do not need to assume that A has finitary operations.

2. **Main theorem.** Our principal result is the partial extension of Theorem 1 to algebras *without* chain condition. As will be seen in §3, this result will contain as special cases many known theorems and some new theorems.

THEOREM 2. *Any algebra A with finitary operations is isomorphic with a subdirect union of subdirectly irreducible algebras.*

PROOF. For any $a\neq b$ of A, consider the set $S(a, b)$ of all congruence relations θ on A, such that $a\not\equiv b \pmod{\theta}$. If T is any linearly ordered subset of $S(a, b)$, the *union* τ of the $\theta\in T$ is defined by the rule

(3) $\qquad x\equiv y \pmod{\tau}$ means $x\equiv y \pmod{\theta}$ for some $\theta\in T$.

It is evident that $a\not\equiv b \pmod{\tau}$, and that *if A has finitary operations,*

[2] This is observed in N. H. McCoy and Deane Montgomery, *A representation of generalized Boolean rings*, Duke Math. J. vol. 3 (1937) p. 46, line 12.

[3] Cf. van der Waerden, *Moderne Algebra*, first ed., vol. 2, p. 36. The unicity theorem on p. 40 does not apply to abstract algebras in general, however.

then τ is a congruence relation. Hence, in the structure lattice $L(A)$ of A, the union of any linearly ordered subset of $S(a, b)$ exists and is in $S(a, b)$. But this is the hypothesis of the "first form" of Zorn's Lemma.[4] The conclusion is that $S(a, b)$ contains a *maximal* element, $\theta_{a,b}$. We next consider $H_{a,b}$, the homomorphic image of A, mod $\theta_{a,b}$.

Every proper congruence relation $\theta > 0$ corresponds[5] to a $\theta' > \theta_{a,b}$; and since $\theta_{a,b}$ is maximal in $S(a, b)$, this implies $a \equiv b \pmod{\theta'}$. Hence the *meet* θ^* of the $\theta > 0$ in $H_{a,b}$, defined by

$$x \equiv y \pmod{\theta^*} \text{ means } x \equiv y \pmod{\theta} \text{ for all } \theta > 0, \tag{4}$$

will satisfy $a \equiv b \pmod{\theta^*}$, and hence $\theta^* > 0$. Hence (cf. §1) $H_{a,b}$ is subdirectly irreducible.

Finally, the meet of *all* $\theta_{a,b}$ is 0, since we have identically $x \not\equiv y \pmod{\theta_{x,y}}$. Hence, by the theorem cited in §1, paragraph 3, A is isomorphic with a subdirect union of the (subdirectly irreducible) $H_{a,b}$, q.e.d.

3. Applications. Theorem 2 has importance mainly because subdirectly irreducible algebras may be specifically described in so many cases.

LEMMA 1. *A weakly irreducible distributive lattice or Boolean algebra must consist of 0 and I alone.*

PROOF FOR DISTRIBUTIVE LATTICES. For any a, the endomorphisms $\theta_a : x \to x \cup a$ and $\theta_a' : x \to x \cap a$ have the property[6] that $\theta_a \cap \theta_a' = 0$, and neither defines the equality relation unless $a = 0$ or $a = I$.

PROOF FOR BOOLEAN ALGEBRAS. Let $x \equiv y \pmod{\theta_a}$ mean $|x - y| \leqq a$ (symmetric difference notation); then $\theta_a \cap \theta_{a'} = 0$, and neither defines the equality relation unless $a = 0$ or $a' = 0$ $(a = I)$.

COROLLARY 1. *Any distributive lattice is isomorphic with a ring of sets.*[7]

COROLLARY 2. *Any Boolean algebra is isomorphic with a field of sets.*[7]

---

[4] Cf. J. W. Tukey, *Convergence and uniformity in topology*, Princeton, 1940, p. 7.

[5] We omit discussing the obvious isomorphism between $L(H_{a,b})$ and the sublattice of $\theta' > \theta_{a,b}$ in $L(A)$.

[6] PROOF. If $x \cup a = y \cup a$ and $x \cap a = y \cap a$, then $x = x \cap (x \cup a) = x \cap (y \cup a) = (x \cap y) \cup (x \cap a) = (y \cap x) \cup (y \cap a) = y \cap (x \cup a) = y \cap (y \cup a) = y$—by the distributive laws. θ_a and θ_a' are endomorphisms.

[7] Corollaries 1–2 are theorems of the author and Stone, respectively. Corollary 3 below is due to McCoy and Montgomery, op. cit. footnote 2. Corollary 4 is due to G. Kothe, *Abstrakte Theorie nichtkommutative Ringe*, Math. Ann. vol. 103 (1930) p. 552; N. H. McCoy *Subrings of infinite direct sums*, Duke Math. J. vol. 4 (1938) pp. 486–494.

LEMMA 2. *A subdirectly irreducible commutative ring R without nilpotent elements is a field.*[8]

PROOF. As in §1, R will have a unique minimal ideal (that is, congruence relation) J. But for any $a \neq 0$ in J, since $aa \neq 0$, $aJ > 0$. Moreover since $(aJ)R = a(JR) \leqq aJ$, aJ is an ideal, $0 < aJ \leqq J$. Consequently $aJ = J$—whence J is a *field* (Huntington's postulates) with unit e. The set $0:e$ of all $x \in R$ such that $ex = 0$ is an ideal, and $(0:e) \cap J = 0$ by what we have just shown; hence $0:e = 0$. But for any $x \in R$, $e(x - ex) = 0$; hence $(x - ex) \in 0:e = 0$, and $x = (x - ex) + ex$, $0 = ex \in J$. We conclude that $R = J$ is a field, q.e.d.

One might infer that by Theorem 2 any commutative ring without nilpotent elements was a subdirect product of fields, but this reasoning would be invalid. It is not necessarily true that a homomorphic image of rings without nilpotent elements is itself without nilpotent elements.

On the other hand, any homomorphic image of a *p-ring* (or commutative ring in which $a^p = a$ for some prime p) is itself a p-ring, and evidently without nilpotent elements, since $a^{p^n} = a$ for all n. Furthermore, a field in which $a^p = a$ can contain only p elements, and must be $GF(p)$ (or the "field" 0).

COROLLARY 3. *Any p-ring is a subdirect union of $GF(p)$, or consists of 0 alone.*[7]

Again, any homomorphic image of a *regular* ring in the sense of von Neumann (or ring in which any a has a "relative inverse" u such that $aua = a$) is itself regular, and evidently without nilpotent elements if commutative (since $a^n u^{n-1} = auau \cdots ua = a \neq 0$).

COROLLARY 4. *Any commutative "regular" ring is a subdirect union of fields.*[7]

If one were interested in obtaining corollaries of Theorem 1, one might show that even a *weakly* irreducible p-ring or regular ring was a field. Again, one might show (van der Waerden, op. cit. p. 32) that, in a weakly irreducible commutative ring satisfying the chain condition, every divisor of zero is nilpotent; this would yield E. Noether's theorem that every commutative ring satisfying the chain condition was a subdirect union of a finite number of primary rings.

Similarly, one can show easily that the only weakly irreducible vector space over a field F is the one-dimensional vector space $V(F; 1)$ (or 0). It follows that any vector space is a subdirect union of one-

[8] This lemma was suggested to the author in conversation by N. H. McCoy.

dimensional vector spaces. Actually (due to the existence of bases) a stronger result is well known.

LEMMA 3. *The only weakly irreducible commutative groups G are the "generalized cyclic" groups: the additive subgroups of the rationals, and those of the rationals* mod *one.*

We omit the proof, which follows easily from the fact that a commutative group with two generators is cyclic unless it contains two disjoint subgroups (the latter hypothesis would make G weakly reducible).

COROLLARY 5. *Any commutative group is a subdirect union of generalized cyclic groups.*

The center of any weakly irreducible hypercentral (alias nilpotent) group H is generalized cyclic (the proof is trivial, granted Lemma 3); the converse also holds if H is finite.[9] Hence we have the following corollary.

COROLLARY 6. *Any hypercentral group is a subdirect union of groups with generalized cyclic centers.*

Further, any weakly irreducible commutative l-group (lattice-ordered group) is known[10] to be *simply* ordered. This yields the following corollary.

COROLLARY 7. *Any commutative l-group is a subdirect union of simply ordered 1-groups.*

One can easily show (although we omit the proof) that any closed element in a closure algebra (in the sense of McKinsey and Tarski[11]) determines a congruence relation, essentially through relativization with respect to the complementary open set. Then from the definition of well-connectedness one obtains the following corollary.

COROLLARY 8. *Any "closure algebra" is a subdirect union of "well-connected" closure algebras.*

HARVARD UNIVERSITY

---

[9] Theorems of Burnside (cf. H. Zassenhaus, *Gruppentheorie*, Teubner, 1937, p. 107, Satz. 11) and P. Hall, *A contribution to the theory of groups of prime-power orders*, Proc. London Math. Soc. vol. 36 (1933) p. 51, Theorem 2.49.

[10] Cf. the author's *Lattice-ordered groups*, Ann. of Math. vol. 43 (1942) p. 319.

[11] *The algebra of topology*, Ann. of Math. vol. 45 (1944) pp. 141–191. The definition of well-connectedness is on p. 147; the concept of relativization is developed on p. 151.

SIMPLE QUASIGROUPS

R. H. BRUCK

Introduction. A. A. Albert [1, II][1] has conjectured that *there exist simple loops of every finite order except order* 4. This conjecture is established in §1 by the construction of what we call *hyperabelian loops*.[2] In §2 other simple loops are constructed, in particular, loops of order $2f+1$ with subloops of order f. The concluding section of the paper is devoted to an investigation of the relationship between non-simple quasigroups (of finite or infinite order) and the loops isotopic to them (Theorem V). Theorems I and V of the paper have elsewhere been announced without proof by the author [2, §10]. For the purposes of §§1 and 2 none of the refinements of the extension theory for loops and quasigroups will be needed [1, 2, 4] and we shall be content in this introductory section with a few remarks sufficient for the proof of Lemma 1 below.

A non-empty set Q of elements $a, b, \cdots$ is said to be a *quasigroup* if (I) *to every ordered pair* $a, b \subset Q$ *there corresponds an element* $ab = c \subset Q$ and (II) *when any two of the symbols* x, y, z *of the equation* $xy = z$ *are assigned as elements of* Q *the third is uniquely determined as an element of* Q. In particular a quasigroup is called a *loop* if it possesses a (unique) unit element. Obviously every group is a loop. The *order* of a quasigroup is, by definition, its cardinal number, finite or transfinite.

If a quasigroup Q is homomorphic to a quasigroup R we may speak of R as a *proper homomorph* of Q if (i) R *is not isomorphic to* Q, (ii) R *is not a group of order one*. Correspondingly a quasigroup is *simple* if it has no proper homomorphs. This definition of simplicity is equivalent to the usual one in the case of groups, as well as to that used for finite quasigroups by G. N. Garrison [4] and to that employed by Albert [1, II] for arbitrary loops.

If a non-simple quasigroup Q has a proper homomorph R we may designate by H_p the set of elements of Q which map into the element p of R under a given homomorphism of Q upon R. Obviously H_p and H_q have common elements if and only if $p = q$. If $pq = r$, let x, y, z designate arbitrary elements of the set H_p, H_q, H_r respectively. Then,

---

Presented to the Society, April 29, 1944; received by the editors April 3, 1944.

[1] Numbers in brackets refer to the references cited at the end of the paper.

[2] The name hyperabelian loop was suggested by the fact that the multiplication table for such a loop F_G is built upon that of a commutative (or abelian) loop. When G is a cyclic group we call F_G hypercyclic.

whenever two of x, y, z are assigned in their respective sets, the third element may be uniquely determined from the equation

$$xy = z \tag{1}$$

as an element of its particular set. If we keep z fixed in (1), we derive a biunique correspondence $x \rightleftarrows y$ between the elements of H_p and H_q from which it follows that *each set H_p has the same cardinal number.* The equation (1) is equivalent to

$$H_p \cdot H_q = H_{pq}, \tag{2}$$

where, as usual, the product on the left of (2) designates the set consisting of all xy with $x \subset H_p$, $y \subset H_q$.

If Q is a loop with unit element e, then R is a loop with a unit element 1. In this case $H \equiv H_1$ is seen from (2) to be a subloop of Q; in fact a *proper* subloop, since $H \neq (e), Q$. Moreover if $x \subset H_p$ then $H_p = xH$. In the light of these remarks the truth of parts (a) and (b) of the following lemma is evident.

LEMMA I. (a) *If, for every proper subloop H of a loop Q, there exist at least two non-identical cosets xH, yH with common elements then Q is simple.*

(b) *If a loop has no proper subloops it is simple.*

(c) *Let G be a loop of finite order g, with a subloop F of order f, where f, g are relatively prime. If every proper subloop of G is a subloop of F then G is simple.*

Part (c) of the lemma requires some additional proof. If G is homomorphic to a proper homomorph R then $H \equiv H_1$ is a proper subloop of G, $H \subset F$. Moreover if H has order h, $h \mid g$, since the disjoint sets H_p exhaust G. Again, F is homomorphic to some loop $S \subset R$, $H \subset F$ consists of all elements of F which map into the identity of S, and so $h \mid f$. Thus h divides $(f, g) = 1$. It follows that $h = 1$, G is isomorphic to R, in contradiction to the hypothesis that R is a proper homomorph of G. Therefore G is simple.

Whether Q is a loop or not we may designate any fixed set H_q as H and put each set H_p in the form $H_p = u \cdot H$. Here $u \subset H_t$, where t is the unique element of R such that $p = tq$. This is the point of view taken by Garrison [4] and later by the author [2].

1. **Hyperabelian loops.** In this section we shall prove the following theorem.

THEOREM I. *Let F be an arbitrary loop of finite order $f \geqq 2$. Let G be a commutative loop of finite order $g \geqq 3$. Then there exists a loop*

$H=F_G$ of order fg with the following properties: (i) *Every proper subloop of H is a subloop of F*; (ii) *H is simple. If $g=2$, H can be constructed with property* (i) *but not with property* (ii).

COROLLARY. *There exist simple loops of every finite order except order* 4.

The loop F_G of the theorem we call a *hyperabelian loop*.[8]

The corollary follows from Theorem I, in view of the facts that every group of prime order is simple (by Lemma I(b)!) and that the only loops of order 4 are the (non-simple) groups.

Before proceeding to the construction of the general loop F_G it will be advantageous to consider the example given by (3) below.

(3) H:

·	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	8
2	2	1	4	5	6	7	8	3
3	3	4	5	6	7	8	2	1
4	4	5	6	7	8	3	1	2
5	5	6	7	8	2	1	3	4
6	6	7	8	3	1	2	4	5
7	7	8	1	2	3	4	5	6
8	8	3	2	1	4	5	6	7

(4) G:

·	E	P	Q	R
E	E	P	Q	R
P	P	Q	R	E
Q	Q	R	E	P
R	R	E	P	Q

Here $F=(1, 2)$ is the cyclic group of order 2, and G is the cyclic group of order 4 given by (4). Thus $f=2$, $g=4$. If $r>f$ is an element of $H=F_G$, then any loop L containing r must contain the unit element 1. But it may be verified from (3), where $(\cdot)$ designates multiplication in H, that

(A) *if $r\cdot s=1$ for r, $s>f$ then either $s\cdot r=2$ or $r\cdot r=2$.*

Thus if L contains $r>f$ we see from (3) that L contains all elements $t\subset H$ with $t>f$. It then follows that $L\supset F$, $L=H$. Thus, in this case, the only proper subloop of H is F itself. But $3\cdot F$ contains the ele-

[8] See the previous footnote.

ments 3, 4 and $4 \cdot F$ contains the elements 4, 5. Hence H is simple, by virtue of Lemma I(a).

In the general case, if $n>4$ is a composite integer, then n has at least one representation of the form $n=fg$ with $f \geqq 2$, $g \geqq 3$. Let G be any arbitrary commutative loop of order g, with elements $E, P, Q, \cdots$, where E is the unit element, and construct the multiplication table for G, as for example in (4). The Cayley square which is formed by deleting the sideline and headline of the multiplication table for G we shall call the *G-square*. Let F be an arbitrary loop of order f (not necessarily commutative), and designate the elements of F by $1, 2, \cdots, f$, where 1 is the unit element of F. Finally, let the elements $1, 2, \cdots, f, f+1, \cdots, n$ designate the elements of the proposed quasigroup $H=F_G$, where 1 is the unit element. In the sideline and headline of the multiplication table for G replace E by a column and row consisting of the numbers $1, 2, \cdots, f$, similarly P by the numbers $f+1, f+2, \cdots, 2f$, and so on, so that the numbers $1, 2, \cdots, n$ are now written in order in sideline and headline.

We shall first give a very simple construction of the multiplication table for $H=F_G$ (in terms of the multiplication for G, modified as above) and then indicate possible generalizations. In the G-square replace that element E which stands in the first row and column of the G-square by the F-square. Thus, as illustrated by (3), the upper left hand corner of the partially constructed multiplication table for H gives the table for F. Now replace each of the remaining elements $P, Q, \cdots$ of the first column of the G-square by a square (not a Cayley square) of f rows and columns, in the following manner: Think of the column formed by $P, Q, \cdots$ as a rectangle consisting of f columns and $n-f$ rows; in the first column write the elements $f+1, f+2, \cdots, n$ in order, and repeat these numbers, in cyclic permutation, in each successive column. (Thus the ith column, $2 \leqq i \leqq f$, consists of the elements $f+i-1, f+i, \cdots, n, f, f+1, \cdots, f+i-2$ written in that order.) Next, replace the element P, wherever it appears in the G-square, by the square which has replaced P in the first column of the G-square, and do the same for Q, and so on, but not for E.

At this stage of the construction it should be clear that 1 has been made the unit element of H and, moreover, that no row or column of the H-square contains the same number more than once. (The second fact has been ensured by use of the Cayley square of a loop G.) Again, if the element E appears below the main diagonal in place (i, j), $1<j<i$, replace E by *any Cayley square $C_{i,j}$ of order f, formed from the elements* $1, 2, \cdots, f$ (for example, by the F-square itself).

Since G is commutative, E must also appear above the main diagonal, in place (j, i); and this time we replace E by the Cayley square $C'_{i,j}$ *derived from* $C_{i,j}$ *by interchanging the numbers* 1 *and* 2. We have thus partially assured the condition (A); in fact *if* $r \cdot s = 1$ *with* $r > f$, $s > f$, *and if* r, s *differ by at least* f, *then* $s \cdot r = 2$. Finally, suppose that E appears on the main diagonal of the G-square, in place (i, i) with $i > 1$, as will happen when G has an element of order two; in this case, replace E by a Cayley square $C_{i,i}$, formed from the elements $1, 2, \cdots, f$, *such that* 2 *appears in every place on the main diagonal, of* $C_{i,i}$.[4] Thus *if* $r \cdot s = 1$ *with* $r > f$, $s > f$, *and if* r, s *differ by less than* f, *then* $r \cdot r = 2$. The multiplication table of H is now complete, and satisfies condition (A); moreover the H-square is a Cayley square, in that each of the numbers $1, 2, \cdots, n$ appears exactly once in each row and column.

In view of condition (A), any subloop S of H which contains an element $r > f$ must contain the element 2. Also, by construction, $n \cdot 2 = f + 1$ and $s \cdot 2 = s + 1$ for $f < s < n$; or (to say the least!)

(B) *the mapping* $r \to r \cdot 2$ *(for* $f < r \leqq n$*) is a permutation of the elements* $f+1, \cdots, n$ *which maps no element into itself.*

It follows that S contains all the elements $f+1, \cdots, n$ and hence every element of H. This much is true even in the excluded case $g = 2$, and hence the constructed $H = F_g$ has property (i) of Theorem I for $g = 2$.

As A. A. Albert has shown [1, II], every loop of order $n = 2f$ $(f > 1)$ which contains a subloop of order f is non-simple. Thus once more we assume $g \geqq 3$. What we have to say here depends essentially upon the construction of the first f columns of the H-square; so we note that

(C) *assuming* $f < r \leqq f(g-1)$, *the first* f *elements of the* r*th row of the* H*-square are, in order,* $r, r+1, \cdots, r+f-1$.

Although we gave the construction by columns, the truth of (C) is readily verified. Now let L be a proper subloop of H, so that $L \subset F$, L has order $l \leqq f$. Then, using (C), we see that the coset $(f+1) \cdot L$ contains the element $f+1$ and, in addition, $l - 1 \geqq 1$ of the elements $f+2, \cdots, 2f$. Let r be the greatest of the elements contained in $(f+1) \cdot L$; hence $f+1 < r \leqq 2f \leqq f(g-1)$. Then, using (C) again, we see that $r \cdot L$ contains r and other elements $t > r$, but not the element $f+1$. Thus $(f+1) \cdot L$ and $r \cdot L$ are distinct cosets with the common element r. By reference to Lemma I(a) we verify that H is a simple loop.

[4] Any Cayley (or latin) square, formed from the elements $1, 2, \cdots, f$, can be converted into a Cayley square with 2 down the main diagonal by suitable permutations of the rows.

Before attempting to generalize the construction of the loop $H=F_G$ we should realize that properties (i) and (ii) of Theorem I contain the essential features of the construction. For (i) we used conditions (A) and (B), and for (ii), condition (C). Now the only important fact about the element 2 used in (A) and (B) is that it is an element of F but not the identity; but to replace 2 by another element would not effect the construction in an essential manner. If 1 is to be the unit element of H we must have $1\cdot i=i\cdot 1=i$ for all i of H, but, aside from this restriction, condition (C) may be relaxed as regards the order in which the first f elements of the rth row appear. Then, if the first f columns of H have been completely filled in, and if the various E's have been replaced by Cayley squares as described above, the rest of the H-square may be filled in any manner, subject only to the obvious restrictions that 1 be the unit element of H and that each of the numbers $1, 2, \cdots, n$ appear exactly once in each row and column of the H-square.

Taking the case that F is the cyclic group of order $f=3$ and that G is the four-group ($g=4$) we give below two simple hypercyclic loops $H=F_G$. That given by (5) was constructed according to the first, or simpler, method, and that given by (6), with the conditions relaxed.[5] When f and g are both fairly large, considerably more freedom of construction is of course possible, without destroying the essential features (i) and (ii) of the hyerabelian loops.

(5)

·	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	2	3	1	5	6	7	8	9	10	11	12	4
3	3	1	2	6	7	8	9	10	11	12	4	5
4	4	5	6	2	3	1	10	11	12	7	8	9
5	5	6	7	1	2	3	11	12	4	8	9	10
6	6	7	8	3	1	2	12	4	5	9	10	11
7	7	8	9	10	11	12	2	3	1	4	5	6
8	8	9	10	11	12	4	1	2	3	5	6	7
9	9	10	11	12	4	5	3	1	2	6	7	8
10	10	11	12	7	8	9	4	5	6	2	3	1
11	11	12	4	8	9	10	5	6	7	1	2	3
12	12	4	5	9	10	11	6	7	8	3	1	2

[5] Note that the differences between (5) and (6) appear in the second and third rows and in the blocks of elements down the main diagonal.

(6)

·	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	2	3	1	6	7	8	9	10	11	12	4	5
3	3	1	2	5	6	7	8	9	10	11	12	4
4	4	5	6	2	1	3	10	11	12	7	8	9
5	5	6	7	3	2	1	11	12	4	8	9	10
6	6	7	8	1	3	2	12	4	5	9	10	11
7	7	8	9	10	11	12	3	1	2	4	5	6
8	8	9	10	11	12	4	2	3	1	5	6	7
9	9	10	11	12	4	5	1	2	3	6	7	8
10	10	11	12	7	8	9	4	5	6	3	2	1
11	11	12	4	8	9	10	5	6	7	1	3	2
12	12	4	5	9	10	11	6	7	8	2	1	3

2. **Other simple loops.** Elsewhere [3, p. 38] the author has announced without proof the following result.

LEMMA II. *Let n be any positive integer. Then an upper bound for the order of any proper sub-quasigroup is $[n/2]$ (the greatest integer in $n/2$). This bound is attained for every n.*

A. A. Albert [1, II] has published a proof of the first statement of this lemma, and has shown moreover that if a quasigroup Q of even order $n=2m$ has a sub-quasigroup of order m then Q is non-simple. We shall now construct the most general *loop* G of odd order $n=2m+1$ which has a subloop F of order $m>1$. (In case $m=1$, G is the cyclic group of order 3.)

THEOREM II. *Let F be an arbitrary loop of finite order $m>1$, consisting of the elements $1, 2, \cdots, m$, with unit element 1. Let G be a system of order $n=2m+1$, consisting of the elements $1, 2, \cdots, n$, closed under a single-valued multiplication, and containing the loop F as a multiplicative subsystem of order m. Let the multiplication table for G be given by*

(7)

A	B_1
C_1	D_1

where it is understood that the sideline and headline each contain the

numbers $1, 2, \cdots, m, m+1, \cdots, n$ *written in natural order. If* A *denotes the Cayley square for* F *then the following conditions are necessary and sufficient in order that* G *be a loop*:

(i) B_1 *must be obtainable by deleting the last row of a Cayley square* B *on the elements* $m+1, \cdots, n$, *of which the first row consists of those elements in natural order.*

(ii) C_1 *must be obtainable by deleting the last column of a Cayley square* C *on the elements* $m+1, \cdots, n$, *of which the first column consists of those elements in natural order.*

(iii) D_1 *must be obtainable from a Cayley square* D *on the elements* $0, 1, 2, \cdots, m$, *in the following manner. Each of the elements* $1, 2, \cdots, m$ *occupies the same position in* D_1 *as in* D. *The latter is partially determined by* B *and* C, *in that if* i (*for* $m+1 \leqq i \leqq n$) *appears in the row numbered* r_i *of the last column of* C *and in the column numbered* c_i *of the last row of* B $(1 \leqq r_i, c_i \leqq m+1)$ *then* 0 *must appear at the intersection of the* r_i*th row and* c_i*th column of* D. *Moreover* i *must appear at the intersection of the* r_i*th row and* c_i*th column of* D_1.

COROLLARY. *For every integer* $m>1$, *there exists a loop* G *of order* $n=2m+1$, *with the following properties*: (a) G *contains an arbitrary loop* F *of order* m. (b) *Every proper subloop of* G *is a subloop of* F. (*By Lemma* I(c), G *is simple.*)

With a few fairly obvious changes Theorem II can be altered to give the construction of the most general *quasigroup* of order $n=2m+1$ with an arbitrary sub-quasigroup of order m.

PROOF OF THEOREM II. Since G is to be a loop with subloop F, the unit element of G must coincide with 1, the unit of F. Hence the first row of B_1 (first column of C_1) must consist of the elements $m+1, \cdots, n$ in natural order. Since A is a Cayley square on the elements $1, 2, \cdots, m$, none of these elements can appear in any row or column of B_1 or of C_1. In particular B_1 consists of m rows and $m+1$ columns formed from the $m+1$ elements $m+1, \cdots, n$. In view of the fact that G is to be a loop, there can be no repetitions in any row or column of B_1; hence each column of B_1 lacks exactly one of the elements $m+1, \cdots, n$, clearly a different one for each column. Thus the missing elements form a uniquely determined final row, which, when added to B_1, yields a Cayley square B on the elements $m+1, \cdots, n$, the first row of which contains the elements in natural order. Thus condition (i) is necessary in order that G be a loop, and similar considerations show the necessity of (ii). Now if the element i $(m+1 \leqq i \leqq n)$ is missing from the r_ith column of B_1, or appears in

the corresponding column of the last row of B, then i must appear somewhere in the r_ith column of D_1. Similarly if the same i is missing from the c_ith row of C_1, it must appear in the c_ith row of D_1. Finally, since this i can appear at most once in each row and column of C_1, it must appear at the intersection of the r_ith row and c_ith column of D_1, and nowhere else in D_1. Moreover, since the elements $1, 2, \cdots, m$ appear nowhere in B and C, each must appear exactly once in each row and column of D_1. Thus if D be obtained from D_1 by replacing each $i>m$ by 0, D must be a Cayley square; it follows that (iii) is necessary in order that G be a loop.

Conversely, let B, C be Cayley squares with first row and final column respectively consisting of the elements $m+1, \cdots, n$ in natural order, let D be a Cayley square on the elements $0, 1, 2, \cdots, m$ with 0 appearing in the places prescribed by (iii), and let B_1, C_1, D_1 be derived from B, C, D as described in the theorem. Then, in the inside of the multiplication table (7), each of the elements $1, 2, \cdots, n$ will appear exactly once in each row and column, inasmuch as A is the given Cayley square for F. Thus G will be a quasigroup, and, in fact, a loop, since it will have unit element 1. Thus the conditions (i), (ii) (iii) are sufficient.

When it comes to the actual problem of constructing G we find it necessary to give further consideration to the relationships between the Cayley squares B, C and D. If B and C are given, subject to the mild restrictions of (i) and (ii), then the positions of the element 0 in D are fixed. But if D' is an arbitrary Cayley square on $0, 1, \cdots, m$ we may permute the rows and columns of D' in order to bring 0 into the prescribed places, and hence there is no difficulty about constructing D. Suppose, on the other hand, that D and B are given arbitrarily. Since D is given, the positions of 0 are fixed in D. Since the last row of B is given, the positions of the elements $m+1, \cdots, n$ in D_1 are fixed, and hence the last column of C is fixed. Thus the first and last columns of C are fixed, and we are faced with the problem of constructing a Cayley square of order $m+1 \geqq 3$ with two given columns. When $m=2$ the remaining column of C is uniquely determined, but whether the construction is always possible for $m>2$ is not obvious.

There is no difficulty in constructing a loop G with the properties of the corollary. For example, let B be the symmetric Cayley square in which successive rows are obtained by cyclic permutations of the first row, and let C be identical with B. Then D will have 0 down the main diagonal, and D_1 will have down the main diagonal the elements $n, m+1, m+2, \cdots, n-1$ in that order. If $i>m$ is any element of G not in F, the elements $i_0=i$, $i_1=i\cdot i$, $i_2=i_1\cdot i_1$, $\cdots$, $i_m=i_{m-1}\cdot i_{m-1}$ will

comprise all elements of G which are not in F, and the subloop generated by i will coincide with G. This establishes the corollary. If F is commutative, the G so constructed will be commutative if and only if D is commutative; but D can be commutative [5, p. 728, or 3, pp. 35–36] if and only if its order $m+1$ is even, or if m is odd, $m=2k+1>1$. Hence we have obtained the following result.

THEOREM III. *Let k be any positive integer, and let F be an arbitrary commutative loop of order $2k+1$. Then there exists a simple commutative loop G of order $4k+3$, with the property that every proper subloop of G is a subloop of F. (In particular F is a subloop of G of maximum possible order.)*

It should be noted conversely that if the G of Theorem II is commutative then C must be the transpose of B, and hence D must be a symmetric Cayley square of order $m+1\geqq 3$ with 0 down the main diagonal. But this can only happen for $m=2k+1\geqq 3$, and so Theorem III gives a "best possible" result of its kind.

We conclude this section with a result due to H. Griffin [5, p. 731].

THEOREM IV. *Let I_n designate a commutative loop of order $n>1$, which contains no proper subloops. Then*

(i) *I_{2^m} does not exist for $m\neq 1$;*

(ii) *I_2, I_3 and I_8 are cyclic groups;*

(iii) *a non-associative I_n exists for every odd integer $n>5$.*

COROLLARY. *When I_n exists, it is simple.*

For the proof of Theorem IV we refer the reader to Miss Griffin's paper,[6] which also contains many other useful constructions. The corollary follows from Lemma I(b).

3. **The extension theory for quasigroups.** In another paper [2, §10] we have defined the extension $P=(H, Q)$ of a set H by a quasigroup Q. It will be convenient to repeat the definition. If H is a set of order m and Q a quasigroup of order n (m, n being finite or transfinite) then P is constructed so as to be a quasigroup of order mn. We suppose that to every pair p, q of elements of Q there has been defined a quasigroup $H_{p,q}$ consisting of the elements of H. No connection is

[6] There seems to be a slight slip in the third sentence of Miss Griffin's construction, which could be corrected by replacing her sentence by these: "Now interchange the elements of the last row, except for the last element, with the principal diagonal elements immediately above them, and alter the last column correspondingly to preserve commutativity. Also change the headline and sideline of the multiplication table to agree respectively with the first row and column of the table."

postulated between the ordered products of a, b in the various $H_{p,q}$, and we denote this product in $H_{p,q}$ by[7] $\phi_{p,q}(a, b)$.

We define $P=(H, Q)$ to be the set product of H and Q, namely the set of all couples (a, p) with $a \subset H$, $p \subset Q$, where $(a, p)=(b, q)$ if and only if $a=b$, $p=q$. The ordered product of $x=(a, p)$ and $y=(b, q)$ is given by $x \cdot y=z=(c, r)$ with

$$c = \phi_{p,q}(a, b), \qquad r = pq. \tag{8}$$

There is now no difficulty in verifying that P is a quasigroup. Moreover, if we designate by H_p the set of all (a, p) with $a \subset H$, then

$$H_p \cdot H_q = H_{pq}, \tag{9}$$

whence it follows that *the set* $\mathfrak{Q}$ *of all sets* H_p *forms a quasigroup isomorphic to* Q, and P *is homomorphic to* $\mathfrak{Q}$. If, conversely, a quasigroup P is homomorphic to a quasigroup S, it is proved in [2] that P is isomorphic to an extension (H, Q) with Q isomorphic to S. Finally, *an extension* (H, Q) *is a loop with unit element* $(e, 1)$ *if and only if*: (i) Q *is a loop with unit* 1; (ii) $H_{1,1}$ *is a loop with unit* e; (iii) e *is a left unit for every* $H_{1,q}$ *and a right unit for every* $H_{p,1}$. Condition (iii) clearly implies (ii). In case P is a loop, A. A. Albert calls $H_{1,1}$ a *normal divisor*, and H may be taken to be identical with $H_{1,1}$.

It is our purpose in this section to supply a proof of the following theorem, previously announced in [2].

THEOREM V. *Let* $P=(H, Q)$ *be a quasigroup extension of a set* H *by a quasigroup* Q. *Then every loop isotopic to* P *is isomorphic to some extension* $P_0=(H_0, Q_0)$ *where* Q_0 *is a loop isotopic to* Q *and* H_0 *is a set (which may be taken to be a loop) of the same order as* H.

COROLLARY. *Every isotope of a simple loop is simple.*

PROOF. If f, g are fixed and x, y arbitrary elements of P, consider the system P_0 with multiplication

$$xoy = xR_g^{-1} \cdot yL_f^{-1}. \tag{10}$$

Here the operators L_f^{-1}, R_g^{-1} designate respectively the inverses of the one-to-one mappings

$$x \to xg = xR_g, \qquad y \to fy = yL_f \tag{11}$$

of P upon itself. It is known [1, 3] that P_0 is a loop and has unit ele-

[7] The only difficulties to be found in the proof of Theorem V of this section are those of notation. It has been found advantageous in this instance to replace certain notations previously used by the author [2, 3] by those recently adopted by Albert [1].

ment $f \cdot g$ (this much may be verified without difficulty from (10) and the definition of a loop), and, moreover, that every loop isotopic to P is isomorphic to a loop P_0 defined for some f and g of P. Hence our proof will consist merely in relating (10) to the fact that $P = (H, Q)$.

If p, q are any elements of Q we define the one-to-one mappings R_p, L_p of Q upon itself by

$$pq = pR_q = qL_p. \tag{12}$$

Similarly, if a, b are any elements of H we define one-to-one mappings $R_a^{p,q}$, $L_b^{p,q}$ of the quasigroup $H_{p,q}$ upon itself by

$$\phi_{p,q}(a, b) = aR_b^{p,q} = bL_a^{p,q}. \tag{13}$$

Assuming that

$$f = (\alpha, u), \qquad g = (\beta, v), \qquad x = (a, p), \qquad y = (b, q), \tag{14}$$

we now may show that

$$xR_g^{-1} = (aV_p^{-1}, pR_v^{-1}), \qquad yL_f^{-1} = (bU_q^{-1}, qL_u^{-1}) \tag{15}$$

where V_p, U_q are one-to-one mappings of H upon itself, defined by

$$\begin{aligned} V_p &= R_\beta^{p',v}, & p' &= pR_v^{-1}, \\ U_q &= L_\alpha^{u,q''}, & q'' &= qL_u^{-1}. \end{aligned} \tag{16}$$

In fact, where g, x are given by (14), the definition of product in $P = (H, Q)$ yields $xR_g = x \cdot g = (\phi_{p,v}(\alpha, \beta), pv) = (aR_\beta^{p,v}, pR_v)$. The inverse mapping R_g^{-1} is uniquely defined; hence if we take it to be given by (15.1) then $x = xR_gR_g^{-1} = (aR_\beta^{p,v}, pR_v)R_g^{-1} = (aR_\beta^{p,v} \cdot V_{pv}^{-1}, p)$. It follows that R_g^{-1} has been correctly defined if and only if

$$V_{pv} = R_\beta^{p,v}.$$

In this last relation we replace p by $pR_v^{-1} = p'$ and obtain (16.1). Similarly we may verify that L_f^{-1} is correctly given by (15.2) if and only if U_q is given by (16.2).

Before proceeding, let us note that V_p depends not only upon p but upon the fixed element $g = (\beta, v)$. A similar remark holds for U_q.

By use of (10), (8), (15) and (16), a direct calculation gives

$$xoy = (\overset{0}{\phi}_{p,q}(a, b), poq) \tag{17}$$

where

$$poq = pR_v^{-1} \cdot qL_u^{-1} \tag{18}$$

and

$$\phi^0_{p,q}(a, b) = \phi_{p',q''}(aV_p^{-1}, bU_q^{-1}). \tag{19}$$

Now (18) defines a loop Q_0, isotopic to Q, and with unit element $u \cdot v \equiv 1$. (Compare the corresponding remarks above about (10) and P.) Moreover, for each fixed pair p, q of elements of Q_0, (19) defines a quasigroup $H^0_{p,q}$ consisting of the elements of H under the multiplication $\phi^0_{p,q}(a, b)$; in fact if p, q are given then p', q'' are determined by (16), and the equation $\phi^0_{p,q}(a, b) = c$, or

$$\phi_{p',q''}(aV_p^{-1}, bU_q^{-1}) = c,$$

enables us to determine any one of a, b, c uniquely once the other two are given. It is indeed evident that *each quasigroup $H^0_{p,q}$ is isotopic to the corresponding quasigroup $H_{p',q''}$*. Since the $H^0_{p,q}$ are quasigroups we see from (17) that P_0 is an extension (H_0, Q_0) of the set $H_0 \equiv H$ by the loop Q_0. But P_0 is itself a loop, and hence the proof of Theorem V is complete.

In view of (10) we know that the unit element of P_0 is

$$(e, 1) \equiv f \cdot g = (\alpha, u)(\beta, v) = (\phi_{u,v}(\alpha, \beta), uv). \tag{20}$$

It would seem of interest to give a direct proof of the facts that $H^0_{1,1}$ is a loop with unit element e, and that e is a left unit for every $H^0_{1,q}$ and a right unit for every $H^0_{p,1}$. Now, from (16), $1' = uv \cdot R_v^{-1} = u$ and $1'' = uv \cdot L_u^{-1} = v$. Moreover $eV_1^{-1} = \phi_{u,v}(\alpha, \beta) \cdot [R^{u,v}_\beta]^{-1} = \alpha$ and $eU_1^{-1} = \phi_{u,v}(\alpha, \beta)[L^{u,v}_\alpha]^{-1} = \beta$. Hence, by (19), (16) and (13),

$$\phi^0_{1,q}(e, b) = \phi_{u,q''}(\alpha, bU_q^{-1}) = bU_q^{-1} \cdot L_\alpha^{u,q''} = b$$

and

$$\phi^0_{p,1}(a, e) = \phi_{p',v}(aV_p^{-1}, \beta) = aV_p^{-1} \cdot R_\beta^{p',v} = a$$

for all a, b. In particular we also have $\phi^0_{1,1}(a, e) = \phi^0_{1,1}(e, a) = a$.

References

1. A. A. Albert, *Quasigroups*. I, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 507–519; II, loc. cit. vol. 55 (1944) pp. 401–419.
2. R. H. Bruck, *Some results in the theory of linear non-associative algebras*, Trans. Amer. Math. Soc. vol. 56 (1944) pp. 141–199.
3. ———, *Some results in the theory of quasigroups*, Trans. Amer. Math. Soc. vol. 55 (1944) pp. 19–52.
4. G. N. Garrison, *Quasi-groups*, Ann. of Math. vol. 41 (1940) pp. 474–484.
5. H. Griffin, *The abelian quasi-group*, Amer. J. Math. vol. 62 (1940) pp. 725–737.

University of Wisconsin

AN APPLICATION OF LATTICE THEORY TO QUASIGROUPS

M. F. SMILEY

The purpose of this note is to show that O. Ore's general formulation of the Jordan-Hölder theorems in partially ordered sets[1] [3] yields the Jordan-Hölder theorems for loops [1] which have recently been obtained by A. A. Albert [2]. We shall assume that the reader is familiar with the papers [1, 2, 3] of O. Ore and of A. A. Albert.

We begin with an examination of certain congruence relations in loops. We are led to characterize the normal divisors of A. A. Albert as those subloops which commute and associate with the elements of the loop. Our proof of the fundamental quadrilateral condition of O. Ore [3] is then based on this characterization.

A *loop* $\mathfrak{G}$ is a quasigroup with an identity element e. For each non-null subset $\mathfrak{H} \subset \mathfrak{G}$ we define a relation H_ρ on $\mathfrak{G}\mathfrak{G}$ as follows:

(1) *If* $x, y \in \mathfrak{G}$, *then* $xH_\rho y$ *in case* $y \in x\mathfrak{H}$.

THEOREM 1. *The relation* H_ρ *is a congruence relation for the loop* $\mathfrak{G}$ *if and only if*

$$\text{(i) } x \in (xh)\mathfrak{H}, \qquad \text{(ii) } (x\mathfrak{H})(y\mathfrak{H}) \subset (xy)\mathfrak{H}$$

for every $x, y \in \mathfrak{G}$ *and* $h \in \mathfrak{H}$.

PROOF. Let H_ρ be a congruence relation for $\mathfrak{G}$. For each $x \in \mathfrak{G}$ and $h \in \mathfrak{H}$ we have $xH_\rho(xh)$ by the definition of H_ρ. Since H_ρ is symmetric we have $(xh)H_\rho x$, $x \in (xh)\mathfrak{H}$. If also $y \in \mathfrak{G}$, $h_1 \in \mathfrak{H}$, then $yH_\rho(hy_1)$, and since H_ρ preserves multiplication we obtain $(xy)H_\rho(xh)(hy_1)$, $(xh)(yh_1) \in (xy)\mathfrak{H}$. Conversely, let $\mathfrak{H}$ satisfy (i) and (ii). If $x \in \mathfrak{G}$, choose $h \in \mathfrak{H}$ and we have $x \in (xh)\mathfrak{H} \subset x\mathfrak{H}$ by (i) and (ii). Thus H_ρ is reflexive. If also $y \in \mathfrak{G}$ and $xH_\rho y$, we have $y = xh_1$, and (i) yields $x \in (xh_1)\mathfrak{H} = y\mathfrak{H}$, $yH_\rho x$. Thus H_ρ is symmetric. If also $z \in \mathfrak{G}$ and $yH_\rho z$, we have $z = yh_2 = (xh_1)h_2 \in x\mathfrak{H}$ by (ii). Thus H_ρ is transitive and is an equivalence relation. The relations $xH_\rho y$, $zH_\rho w$ yield $(xz)H_\rho(yw)$ by (ii) and we conclude that H_ρ is a congruence relation for $\mathfrak{G}$.

Remark 1. If R is a congruence relation for a loop $\mathfrak{G}$, then the subset $\mathfrak{R} \equiv [x; xRe]$ is a *subloop* of $\mathfrak{G}$. For clearly $e \in \mathfrak{R}$ and $\mathfrak{R}$ is closed with

Presented to the Society, April 29, 1944; received by the editors April 8, 1944. The opinions and assertions contained in this paper are the private ones of the author and are not to be construed as official or reflecting the views of the United States Navy Department or of the naval service at large.

[1] Numbers enclosed in brackets denote the references given at the end of the paper.

respect to multiplication. If we have $r_1x=r_2$, then eRr_1, xRx give $xR(r_1x)$, xRr_2, r_2Re, xRe, $x\in\mathfrak{R}$. Similarly $xr_1=r_2$ implies $x\in\mathfrak{R}$ and $\mathfrak{R}$ is a subloop of $\mathfrak{G}$. With $R=H_\mathfrak{H}$, we obtain $\mathfrak{R}=\mathfrak{H}$ and thus (i) and (ii) imply that $\mathfrak{H}$ is a subloop of $\mathfrak{G}$. (For groups $\mathfrak{G}$, every congruence relation R is of the form $H_\mathfrak{H}$ for suitable $\mathfrak{H}\subset\mathfrak{G}$.)

Remark 2. We note the following consequences of (i) and (ii):

$$x(y\mathfrak{H}) \subset (xy)\mathfrak{H}, \tag{2}$$

$$(x\mathfrak{H})y \subset (xy)\mathfrak{H}, \tag{3}$$

$$x(\mathfrak{H}y) \subset (xy)\mathfrak{H}, \tag{4}$$

$$\mathfrak{H}x \subset x\mathfrak{H}. \tag{5}$$

For *finite* $\mathfrak{H}$ each of these inequalities[2] is an *equality*. To see this for (2), note that for fixed $x, y\in\mathfrak{G}$ we have a single-valued mapping $h\to h^*$ defined by $x(yh)=(xy)h^*$. This mapping is one-to-one since $x(yh_1)=x(yh_2)$ implies $h_1=h_2$. For a finite $\mathfrak{H}$, we conclude that the set of images $[h^*]=\mathfrak{H}$, and consequently that $(xy)\mathfrak{H}=x(y\mathfrak{H})$. The same reasoning applies to the inequalities (3), (4), and (5). It is now easy to see that if $\mathfrak{H}$ is a *finite* subloop of $\mathfrak{G}$, then $H_\mathfrak{H}$ is a congruence relation for $\mathfrak{G}$ if and only if the set $\mathfrak{H}$ commutes and associates with the elements of $\mathfrak{G}$ in the sense that

$$x\mathfrak{H} = \mathfrak{H}x, \tag{6}$$

$$(xy)\mathfrak{H} = x(y\mathfrak{H}), \qquad (x\mathfrak{H})y = x(\mathfrak{H}y), \qquad (\mathfrak{H}x)y = \mathfrak{H}(xy) \tag{7}$$

for every $x, y\in\mathfrak{G}$.

THEOREM 2. *A subloop $\mathfrak{H}$ of a loop $\mathfrak{G}$ satisfies* (6) *and* (7) *if and only if $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$ in the sense of A. A. Albert.*

PROOF. Consider a subloop $\mathfrak{H}$ of a loop $\mathfrak{G}$ which satisfies (6) and (7). The permutation group $\mathfrak{H}_r$ generated by the right multiplication of $\mathfrak{G}$ gives rise to the cosets $x\mathfrak{H}_r$ of Albert. We first show that the cosets $x\mathfrak{H}_r=x\mathfrak{H}$. Clearly $x\mathfrak{H}_r\supset x\mathfrak{H}$. The elements of $\mathfrak{H}_r$ are finite products of permutations of the form R_h and R_h^{-1} for $h\in\mathfrak{H}$. If $y=xR_h^{-1}$, then $yh=x$, $y=y(hh_1)=(yh)h_2=xh_2$, $y=xR_{h_2}$. It follows that each element of $x\mathfrak{H}_r$ may be written in the form $xh\in x\mathfrak{H}$. We next note that

$$(x\mathfrak{H})(y\mathfrak{H}) = (xy)\mathfrak{H} \tag{8}$$

is an immediate consequence of (6) and (7). To see that the cosets $x\mathfrak{H}$ form a loop we now consider the equations

[2] This result is due to G. N. Garrison. See §4 of his fundamental paper, *Quasigroups*, Ann. of Math. vol. 41 (1940) pp. 474–487.

$$(9) \qquad (x\mathfrak{H})(w\mathfrak{H}) = y\mathfrak{H}, \qquad (z\mathfrak{H})(x\mathfrak{H}) = y\mathfrak{H}.$$

Solutions of (9) are provided by the solutions of $xw=y$ and $zx=y$ for w and z. The uniqueness of the solutions $w\mathfrak{H}$ and $z\mathfrak{H}$ of (9) is readily verified. We have proved that $\mathfrak{G}/\mathfrak{H}$ is a loop and hence that $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$ in the sense of A. A. Albert.

To prove the converse it suffices to use the representation $\mathfrak{H}=e\Gamma$ furnished by Albert's Theorem 3 [1]. Using this theorem, the verifications of (6) and (7) are quite simple and we shall omit them.

Remark 3. The equations (6) and (7) may be replaced by

$$(10) \qquad (x\mathfrak{H})y = x(y\mathfrak{H}),$$

$$(11) \qquad (xy)\mathfrak{H} = x(y\mathfrak{H})$$

for every $x, y\in\mathfrak{G}$. To see this, set $x=e$ in (10) to obtain (6).

THEOREM 3. *The intersection $\mathfrak{H}$ of a system $(\mathfrak{H}_\alpha;\ \alpha\in\Omega)$ of normal divisors of a loop $\mathfrak{G}$ is a normal divisor of $\mathfrak{G}$.*

PROOF. Clearly $\mathfrak{H}$ is a subloop of $\mathfrak{G}$. We shall prove $\mathfrak{H}x\subset x\mathfrak{H}$ for every $x\in\mathfrak{G}$. If $h\in\mathfrak{H}$, $\alpha\in\Omega$, we have $hx=xh_\alpha$ for some $h_\alpha\in\mathfrak{H}_\alpha$. The left cancellation law then shows that the set $[h_\alpha;\ \alpha\in\Omega]$ is singular and its sole member is an element of $\mathfrak{H}$. The remainder of the proof consists in a repetition of this argument which proves (6) and (7).

COROLLARY. *The set of all normal divisors of a loop $\mathfrak{G}$, when ordered by set inclusion, forms a complete lattice.*

PROOF. It suffices to remark that $\mathfrak{G}$ is a normal divisor of $\mathfrak{G}$. We shall use $\cup$ and $\cap$ to denote the lattice operations of this lattice.

THEOREM 4. *If $\mathfrak{H}$ and $\mathfrak{K}$ are normal divisors of a loop $\mathfrak{G}$, then $\mathfrak{H}\mathfrak{K}$ is a normal divisor of $\mathfrak{G}$ and $\mathfrak{H}\mathfrak{K}=\mathfrak{H}\cup\mathfrak{K}$.*

PROOF. Clearly $e\in\mathfrak{H}\mathfrak{K}$ and $\mathfrak{H}\mathfrak{K}$ is closed with respect to multiplication. Now if $x(hk)=h_1k_1$, we use (7) to get $(xh)k_2=h_1k_1$. Two applications of (i) and use of (6) and (7) then give $x=h_3k_3$. We compute $x(\mathfrak{H}\mathfrak{K})=(x\mathfrak{H})\mathfrak{K}=(\mathfrak{H}x)\mathfrak{K}=\mathfrak{H}(x\mathfrak{K})=\mathfrak{H}(\mathfrak{K}x)=(\mathfrak{H}\mathfrak{K})x$. Thus $\mathfrak{H}\mathfrak{K}$ is a subloop of $\mathfrak{G}$ satisfying (6). We omit the simple verification of (7). Since every normal divisor of $\mathfrak{G}$ which contains $\mathfrak{H}$ and $\mathfrak{K}$ must contain $\mathfrak{H}\mathfrak{K}$, we conclude that $\mathfrak{H}\mathfrak{K}=\mathfrak{H}\cup\mathfrak{K}$.

COROLLARY. *The lattice of normal divisors of a loop $\mathfrak{G}$ is modular.*

THEOREM 5 (QUADRILATERAL CONDITION). *If $\mathfrak{H}$ and $\mathfrak{K}$ are maximal normal divisors of a loop $\mathfrak{G}$, then $\mathfrak{H}\cap\mathfrak{K}$ is a maximal normal divisor of the loop $\mathfrak{H}$, and $\mathfrak{G}/\mathfrak{K}\cong\mathfrak{H}/\mathfrak{H}\cap\mathfrak{K}$.*

PROOF. Suppose on the contrary that $\mathfrak{L}$ is a normal divisor of the loop $\mathfrak{H}$ such that $\mathfrak{H} > \mathfrak{L} > \mathfrak{H} \cap \mathfrak{K}$. We shall prove that $\mathfrak{L}\mathfrak{K}$ is a normal divisor of $\mathfrak{G}$. We note that $\mathfrak{G} = \mathfrak{H}\mathfrak{K}$ and to verify (10) we consider the equation

$$[(hk)(h_1k_1)](h_2k_2) = (hk)[(h_2k_2)(h_3k_3)]$$

which we may write, using (8), (i), and (7) for $\mathfrak{K}$, as

$$(hh_1)h_2 = [h(h_2h_3)]k_4. \tag{12}$$

Now if $h_1 \in \mathfrak{L}$, we may use (10) for $\mathfrak{L}$ to get $h(h_2l) = [h(h_2h_3)]k_4$. We now use (7) for $\mathfrak{K}$, cancel h, use (7) for $\mathfrak{K}$ again and cancel h_2 to obtain $l = h_3k_5$. It follows that $k_5 \in \mathfrak{H}$, and hence that $k_5 \in \mathfrak{H} \cap \mathfrak{K} \subset \mathfrak{L}$. Consequently $h_3 \in \mathfrak{L}$ and we have proved $[x(\mathfrak{L}\mathfrak{K})]y \subset x[y(\mathfrak{L}\mathfrak{K})]$ for every x, $y \in \mathfrak{G}$. On the other hand, if $h_3 \in \mathfrak{L}$, we may use (10) for $\mathfrak{L}$ to get $(hh_1)h_2 = [(hl)h_2]k_4$. We now use (6) and (7) for $\mathfrak{K}$ and cancel h_2, use (6) and (7) for $\mathfrak{K}$ and cancel h to obtain $h_1 = lk_5$ from which $h_1 \in \mathfrak{L}$ follows as before. This completes the verification of (10). We apply the same reasoning to verify (11). We write

$$[(hk)(h_1k_1)](h_2k_2) = (hk)[(h_1k_1)(h_3k_3)]$$

as $(hh_1)h_2 = [h(h_1h_3)]k_4$. Then if $h_3 \in \mathfrak{L}$, we have $h(h_1l) = h[(h_1h_3)k_5]$ and we easily obtain $h_3 \in \mathfrak{L}$. Also, if $h_3 \in \mathfrak{L}$, we have $(hh_1)h_2 = (hh_1)(lk_5)$, $h_2 = lk_5$, $h_2 \in \mathfrak{L}$. Thus (11) holds for $\mathfrak{L}\mathfrak{K}$. To show that $\mathfrak{L}\mathfrak{K}$ is a subloop of $\mathfrak{G}$ it now suffices to note that $e \in \mathfrak{L}\mathfrak{K}$, $(\mathfrak{L}\mathfrak{K})(\mathfrak{L}\mathfrak{K}) = \mathfrak{L}\mathfrak{K}$, and to prove that if $(hk)(l_1k_1) = l_2k_2$, then $h \in \mathfrak{L}$. But we may write this equation as $hl_1 = l_2k_4$ and get $h = h(l_1l_2) = (hl_1)l_4 = l_5k_5$. As before, $k_5 \in \mathfrak{L}$, $h \in \mathfrak{L}$. We have proved that $\mathfrak{L}\mathfrak{K}$ is a normal divisor of $\mathfrak{G}$. Note also that $\mathfrak{G} > \mathfrak{L}\mathfrak{K} > \mathfrak{K}$, since if $h \in \mathfrak{H}$ and h does not belong to $\mathfrak{L}$, then $hk = lk_1$ implies $h = lk_2$, $k_2 \in \mathfrak{L}$, $h \in \mathfrak{L}$, a contradiction, and $\mathfrak{L} > \mathfrak{H} \cap \mathfrak{K}$. Our assumption of the existence of a normal divisor $\mathfrak{L}$ of $\mathfrak{H}$ satisfying $\mathfrak{H} > \mathfrak{L} > \mathfrak{H} \cap \mathfrak{K}$ has led to a contradiction of the assumption that $\mathfrak{K}$ is a maximal normal divisor of $\mathfrak{G}$. We conclude that $\mathfrak{H} \cap \mathfrak{K}$ is a maximal normal divisor of $\mathfrak{H}$.

To establish the isomorphism we map the cosets $x\mathfrak{K}$ of $\mathfrak{G}/\mathfrak{K}$ onto the cosets $x(\mathfrak{H} \cap \mathfrak{K})$ of $\mathfrak{H}/\mathfrak{H} \cap \mathfrak{K}$. This mapping exhausts the set $\mathfrak{H}/\mathfrak{H} \cap \mathfrak{K}$ and is one-to-one. For, if $x(\mathfrak{H} \cap \mathfrak{K}) = y(\mathfrak{H} \cap \mathfrak{K})$, then $xk = yu$ gives $x = (yu)k_1 = yk_2$, $uk_3 = k_2$, $u \in \mathfrak{K}$, $x\mathfrak{K} \subset y\mathfrak{K}$, and symmetry yields $x\mathfrak{K} = y\mathfrak{K}$. That this mapping preserves multiplication in the two loops $\mathfrak{G}/\mathfrak{H}$ and $\mathfrak{H}/\mathfrak{H} \cap \mathfrak{K}$ is evident. The proof is complete.

Theorem 5 is the main tool which is needed to apply Ore's result [3] to the partially ordered set of subloops of a loop $\mathfrak{G}$ which occur in some composition series of $\mathfrak{G}$.

REFERENCES

1. A. A. Albert, *Quasigroups*. I, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 507–519.
2. ———, *Quasigroups*. II, Trans. Amer. Math. Soc. vol. 55 (1944) pp. 401–419.
3. O. Ore, *Chains in partially ordered sets*, Bull. Amer. Math. Soc. vol. 49 (1943) pp. 558–566.

UNITED STATES NAVAL ACADEMY, POSTGRADUATE SCHOOL

TWO ELEMENT GENERATION OF A SEPARABLE ALGEBRA

A. A. ALBERT

The *minimum rank* of an algebra A over a field F is defined to be the least number $r=r(A)$ of elements $x_1, \cdots, x_r$ such that A is the set of all polynomials in $x_1, \cdots, x_r$ with coefficients in F. In what follows we shall assume that A is an *associative algebra of finite order* over an *infinite* field F.

It is well known that $r(A)=1$ if A is a separable field over F and that $r(A)=2$ if A is a total matric[1] algebra over F. Over fourteen years ago I obtained but did not publish the result that $r(A)=2$ if A is a central division algebra over F. The purpose of this note is to provide a brief proof of the generalization which states that if A is any *separable* algebra over F then $r(A)=1$ or 2 according as A is or is not commutative.

We observe first that a commutative separable[2] algebra Z is a direct sum of separable fields and that there exists a scalar extension K over F such that Z_K has a basis $e_1, \cdots, e_n$ over F for pairwise orthogonal idempotents e_i. If $u_1, \cdots, u_n$ is a basis of Z over F and $x=a_1u_1+\cdots+a_nu_n$ the powers x^i have the form

$$x^i = \sum_{j=1}^{n} b_{ij}u_j \qquad (i = 1, \cdots, n),$$

where the determinant

$$d(a_1, \cdots, a_n) = |b_{ij}|$$

is a polynomial in the parameters $a_1, \cdots, a_n$. If $c_1, \cdots, c_n$ are any

Received by the editors April 13, 1944.

[1] See page 95 of my *Modern higher algebra*.

[2] The definition of a separable algebra given below reduces to a direct sum of fields in the commutative case. When F is nonmodular the concept of semisimple algebra and separable algebra coincide.

distinct elements of K the quantity $x_0=c_1e_1+\cdots+c_ne_n$ is known[1] to generate the diagonal algebra Z_K, that is $Z_K=F[x_0]$ has a basis $x_0, x_0^2, \cdots, x_0^n$. If we express the quantities $e_1, \cdots, e_n$ linearly in terms of $u_1, \cdots, u_n$ we see that x_0 is a value of x for values a_{i0} of the a_i in K. The linear independence of $x_0, x_0^2, \cdots, x_0^n$ implies that $d(a_{10}, \cdots, a_{n0})\neq 0$. Then $d(a_1, \cdots, a_n)$ is not identically zero and thus there exists a quantity x in Z such that Z has a basis $x, x^2, \cdots, x^n$ over F, $Z=F(x)$, $r(Z)=1$.

An algebra A is called a separable[2] algebra if A is a direct sum of simple components A_k such that the center of every A_k is a separable field over F. If x and y are in A we define

$$F[x, y]$$

to mean the set of all polynomials

$$\sum_{i=1,\cdots,m}^{j=1,\cdots,q} a_{ij}x^iy^j \qquad (a_{ij} \text{ in } F).$$

Only a finite number of the power products x^iy^j are linearly independent and each $F[x, y]$ is a linear subspace of A, m and q may be be selected so that $F[x, y]$ has order mq over F.

A separable algebra has a unity quantity e and if $A=F[x]$ then A has a basis $x^0=e, x, \cdots, x^{n-1}$ over F, $A=F[x, e]$. Also A is commutative. If A is not commutative and $A=F[x, y]$ then $e=x[f(x, y)]y$ and thus x and y must be nonsingular. Note then that A has a basis of power products x^iy^j where $i=0, \cdots, m-1$ and $j=0, \cdots, q-1$. We use these results in the proof of our principal

THEOREM. *If A is a separable algebra which is not commutative then $r(A)=2$, $A=F[x, y]$ for nonsingular elements x and y such that $F[x]$ is separable.*

We first study the case where A is the direct product of a total matric algebra M of degree g and a division algebra D of degrees s over a separable center C over F. It is well known[3] that D contains a maximal separable subfield $W=C[x_0]$ of degree s over C and that $W=F[x_0]$. The algebra $Q=(e_{11}, \cdots, e_{gg})$, whose basis consists of a set of primitive idempotents of M whose sum is its unity element e, has the property that $Z\neq Q\times W$ is separable and commutative, and so $Z=F[x]$. If K is a scalar splitting field over C of D the algebra Z_K contains $n=gs$ primitive pairwise orthogonal idempotents whose sum

[3] See Theorem 4.18 of my *Structure of algebras*, Amer. Math. Soc. Colloquium Publications, vol. 24, New York, 1939.

is the unity element e of the total matric algebra A_K of degree n over K. Also $Z=C[x]$, $Z_K=K[x]$ and it is known[1] that there exists a quantity y_0 in A_K such that

$$A_K = K[x, y_0],$$

that is, the power products $x^i y_0^j$ taken for $i, j=1, \cdots, n$ are linearly independent in K. If $p=n^2$ and $u_1, \cdots, u_p$ are a basis of A over C we may write $y=a_1u_1+\cdots+a_pu_p$ and express the powers x^iy^j in the form

$$z_k = x^iy^j = \sum_{h=1} b_{kh}u_h$$
$$(k = i + jn - n;\ i, j = 1, \cdots, n),$$

for b_{kh} in F. The determinant $d(a_1, \cdots, a_p)=|b_{kh}|$ is a polynomial in $a_1, \cdots, a_p$ with coefficients in C which is not identically zero since it is not zero for values $a_{10}, \cdots, a_{p0}$ which define y_0. It follows that $A=C[x, y]$. But $C[x]=F[x]$ so that $A=F[x, y]$.

We finally consider a separable algebra A which is the direct sum of simple components $A_1, \cdots, A_t$. By the proofs above every component $A_k=F[x_k, y_k]$, where y_k is the unity quantity e_k of A_k when A_k is commutative, $Z_k=F[x_k]$ is separable. The algebra Z which is the direct sum of $Z_1, \cdots, Z_t$ is a commutative separable algebra and so $Z=F[x]$. Let $y=y_1+\cdots+y_t$. Since $F[x]$ contains every x_k the linear space $F[x, y]$ contains $x_k^iy^j=x_k^iy_k^j$. For $x_k^i=x_k^ie_k$ and $e_ky^j=(e_ky)^j=y_k^j$. It follows that $F[x, y]$ contains every A_k and that $F[x, y]=A$.

THE UNIVERSITY OF CHICAGO

LIMITS FOR THE CHARACTERISTIC ROOTS OF A MATRIX[1]

A. B. FARNELL

Let A be a square matrix of order n with complex numbers as elements. The equation $|\lambda I - A| = 0$ is called the characteristic equation of the matrix A, and the roots λ_i, the characteristic roots of the matrix A. Although it is not possible to make any definite statements regarding the nature of the characteristic roots for the general matrix, several authors have given upper limits to the roots. In 1900, Bendixson [1][2] obtained upper limits for the real and imaginary parts of the characteristic roots of a real matrix. In 1902, Hirsch [5] extended these results to matrices with complex numbers as elements. A limit was also given by Bromwich [2] in 1904. These limits were further refined by Browne [3] in 1930, and by Parker [7] in 1937.

In 1918, Toeplitz [8], using the results of Bendixson and Hirsch, studied the algebraic form $(Ax|x) \equiv \sum_{r,s} a_{rs} x_s \bar{x}_r$ corresponding to the matrix A, where by hypothesis the algebraic form $(x|x) = \sum_{r=1}^{n} x_r \bar{x}_r$ has the value unity. By using the fact that A may be decomposed uniquely in the form $A = B + iC$, where B and C are Hermitian, he showed that the totality of values which the algebraic form assumes lie within a rectangle with sides parallel to the real and imaginary axes. He further showed that this field of values is bounded by a convex curve. Hausdorff [4] showed that the field of values is connected, bounded, closed, and convex. In 1932, Murnaghan [6], using the fact that $(Ax|x)$ has values invariant under unitary transformations of A, showed that for normal matrices (that is, matrices which can be transformed unitarily into diagonal form) the field of values is a convex polygon. For non-normal matrices, he showed that in the general case the characteristic roots of the matrix are the foci of the curve bounding the field of values. Wintner had previously remarked that for $n = 2$ the curve is an ellipse. The first three theorems below give radii of circles within which not merely the characteristic roots but also the entire field of values lies.

Let A' and $\overline{A}$ denote the transpose and conjugate, respectively, of the matrix A, and write

$$B = (A + \overline{A}')/2, \qquad C = (A - \overline{A}')/2i.$$

Received by the editors March 10, 1944, and, in revised form, April 10, 1944.

[1] Partially extracted from a dissertation, University of California, 1944.

[2] Numbers in brackets refer to the Bibliography at the end of the paper.

It is evident that B and C are Hermitian. The theorem given by Parker may be stated as follows:

PARKER'S THEOREM. *If A is any square matrix, and if $2Sr$, $2S'r$, $2S''r$ are the sums of the absolute values of the elements in the rth row and the absolute values of the elements in the rth column of A, B, and C, respectively, and if S, S', S'' are the greatest of the Sr, $S'r$, $S''r$, respectively, then for any characteristic root, $\lambda=\alpha+i\beta$, of A, we have*

$$|\lambda| \leqq S, \qquad |\alpha| \leqq S', \qquad |\beta| \leqq S''.$$

If $\lambda=\alpha+i\beta$ is a characteristic root of a matrix $A=(a_{rs})$ of order n, there exists a set of numbers $(x_1, x_2, \cdots, x_n)$ such that

$$\sum_{r=1}^{n} x_r\bar{x}_r = 1,$$

which satisfy the relations

$$\lambda x_r = \sum_s a_{rs}x_s \qquad (r = 1, 2, \cdots, n). \tag{1}$$

If we multiply the rth equation in (1) by $\bar{x}_r$, and sum as to r, we obtain

$$\lambda = \sum_{r,s} a_{rs}\bar{x}_r x_s. \tag{2}$$

The numbers $\sum a_{rs}\bar{x}_r x_s$, where $\sum x_r\bar{x}_r$ has the value unity, comprise the field of values of A. Hence (2) shows that the characteristic roots lie within the field of values. In the following, λ will be considered to be any value in this field, the conclusions holding a fortiori where λ is a characteristic root.

Taking conjugates of both sides of (2), and interchanging subscripts, we have

$$\bar{\lambda} = \sum_{r,s} \bar{a}_{sr}\bar{x}_r x_s. \tag{3}$$

From (2) and (3), by addition and subtraction, it follows that

$$\begin{aligned} \alpha &= \sum_{r,s} b_{rs}\bar{x}_r x_s, \\ \beta &= \sum_{r,s} c_{rs}\bar{x}_r x_s. \end{aligned} \tag{4}$$

From the relations (2) and (4), upper limits for $|\lambda|$, $|\alpha|$, and $|\beta|$ may be obtained. Since these relations are identical in form, it is sufficient to carry through the computations for one of them. Further-

more, the theorems will be stated briefly since the context will be evident.

Several results will now be established. The best upper limit may be given by one in some cases, by another in others.

THEOREM 1. *Let* $R_r=\sum_s|a_{rs}|$, $T_s=\sum_r|a_{rs}|$, $R=\max(R_r)$, $T=\max(T_s)$. *Then*

$$|\lambda| \leqq (RT)^{1/2}.$$

Let $\xi_r=|x_r|$, so that

$$\sum_r \xi_r^2 = 1.$$

If we take the absolute value of (2), we have

$$|\lambda| \leqq \sum_{r,s} |a_{rs}|\,\xi_r\xi_s. \tag{5}$$

Rewriting $|a_{rs}|\xi_r\xi_s$ as $|a_{rs}|^{1/2}\xi_r\cdot|a_{rs}|^{1/2}\xi_s$, and applying Schwarz's inequality, it follows that

$$\begin{aligned}
|\lambda| &\leqq \left(\sum_{r,s}|a_{rs}|\,\xi_r^2\right)^{1/2}\left(\sum_{r,s}|a_{rs}|\,\xi_s^2\right)^{1/2} \\
&= \left(\sum_r R_r\xi_r^2\right)^{1/2}\left(\sum_s T_s\xi_s^2\right)^{1/2} \\
&\leqq \left(R\sum_r \xi_r^2\right)^{1/2}\left(T\sum_s \xi_s^2\right)^{1/2} \\
&= (RT)^{1/2}.
\end{aligned}$$

THEOREM 2. *If* λ *lies in the field of values of* A, *then,*

$$|\lambda| \leqq \left(\sum_{r,s}|a_{rs}|^2\right)^{1/2}$$

Rewriting $|a_{rs}|\xi_r\xi_s$ in (5) as $|a_{rs}|\cdot\xi_r\xi_s$, and applying Schwarz's inequality, we have

$$\begin{aligned}
|\lambda| &\leqq \left(\sum_{r,s}|a_{rs}|^2\right)^{1/2}\left(\sum_{r,s}\xi_r^2\xi_s^2\right)^{1/2} \\
&= \left(\sum_{r,s}|a_{rs}|^2\right)^{1/2}\left(\left(\sum_r \xi_r^2\right)^2\right)^{1/2} \\
&= \left(\sum_{r,s}|a_{rs}|^2\right)^{1/2}.
\end{aligned}$$

THEOREM 3. *If λ lies in the field of values of A, then $|\lambda| \leqq M$, where M is defined by (7) below.*

From (2),

$$
\begin{aligned}
|\lambda| &= \left| \sum_{r,s} a_{rs} \bar{x}_r x_s \right| \\
&\leqq \frac{1}{n} \sum_r |a_{rr}| + \frac{1}{n} \sum_s \sum_{r<s} (|a_{rr}| - |a_{ss}|)(\xi_r^2 - \xi_s^2) \\
&\quad + 2 \sum_s \sum_{r<s} |a_{rs}| \xi_r \xi_s \\
&= \frac{1}{n} \sum_r |a_{rr}| + \sum_s \sum_{r<s} \frac{(n-1)^{1/2}}{n} (|a_{rr}| - |a_{ss}|) \cdot \frac{\xi_r^2 - \xi_s^2}{(n-1)^{1/2}} \\
&\quad + 2 \sum_s \sum_{r<s} \left(\frac{n-1}{2n} \right)^{1/2} |a_{rs}| \cdot \left(\frac{2n}{n-1} \right)^{1/2} \xi_r \xi_s \\
&\leqq \frac{1}{n} \sum_r |a_{rr}| + \left\{ \frac{n-1}{n^2} \sum_s \sum_{r<s} (|a_{rr}| - |a_{ss}|)^2 \right. \\
&\quad \left. + \frac{n-1}{2n} \sum_s \sum_{r<s} (|a_{rs}| + |a_{sr}|)^2 \right\}^{1/2}
\end{aligned} \tag{6}
$$

by Schwarz's inequality, and the fact that $\sum_r \xi_r^2 = 1$. That is, $|\lambda| \leqq M$, where

$$
M = \frac{1}{n} \left[\sum_r |a_{rr}| + \left\{ (n-1) \sum_s \sum_{r<s} (|a_{rr}| - |a_{ss}|)^2 + \frac{n(n-1)}{2} \sum_s \sum_{r<s} (|a_{rs}| + |a_{sr}|)^2 \right\}^{1/2} \right]. \tag{7}
$$

For $n=2$, M is the maximum value of the right-hand side of (6), and hence the minimum for this type of upper limit. For $n>2$, the form of the maximum for the right-hand side of (6) is too complicated for practical use.

THEOREM 4. *Let* $U_r = \sum_s |a_{rs}|^2$, $V_s = \sum_r |a_{rs}|^2$. *Then* $|\lambda| \leqq (\sum_r (U_r V_r)^{1/2})^{1/2}$.

From (1),

$$\lambda x_r = \sum_s a_{rs} x_s.$$

Multiplying each by λ, and replacing λx_r from (1), we have

$$\lambda^2 x_1 = a_{11}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + \cdots + a_{1n}(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

$$\lambda^2 x_n = a_{n1}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + \cdots + a_{nn}(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n).$$

Multiplying the rth equation by $\bar{x}_r$, adding the n equations, and taking the absolute value of both sides, we get

$$\begin{aligned}\lambda^2 \leqq\ & (|a_{11}|\xi_1 + |a_{21}|\xi_2 + \cdots + |a_{n1}|\xi_n)(|a_{11}|\xi_1 + |a_{12}|\xi_2 \\ & + \cdots + |a_{1n}|\xi_n) + \cdots + (|a_{1n}|\xi_1 + |a_{2n}|\xi_2 + \cdots \\ & + |a_{nn}|\xi_n)(|a_{n1}|\xi_1 + |a_{n2}|\xi_2 + \cdots + |a_{nn}|\xi_n) \\ \leqq\ & \sum_r (U_r V_r)^{1/2}, \end{aligned} \tag{8}$$

since, by Schwarz's inequality, $\sum_r C_r\xi_r \leqq (\sum_r C_r^2)^{1/2}(\sum_r \xi_r^2)^{1/2} = (\sum_r C_r^2)^{1/2}$.

THEOREM 5. *For $n=2$, $|\lambda| \leqq (RT)^{1/2}$, where $RT = \max(R_rT_r)$.*

From (8),

$$\begin{aligned}\lambda^2 \leqq\ & (|a_{11}|\xi_1 + |a_{12}|\xi_2)(|a_{11}|\xi_1 + |a_{21}|\xi_2) \\ & + (|a_{12}|\xi_1 + |a_{22}|\xi_2)(|a_{21}|\xi_1 + |a_{22}|\xi_2).\end{aligned} \tag{9}$$

Here, for example, $R_2T_2 \leqq R_1T_1$ implies $|a_{22}| \leqq |a_{11}|$. Then since $\xi_1\xi_2 \leqq 1/2$, (9) gives,

$$\lambda^2 \leqq R_1T_1.$$

The probability of the truth of this theorem was suggested to me by W. V. Parker. Although no contradictions have been found, no method of proof has been discovered for $n>2$.

The effectiveness of the above theorems can sometimes be improved by using the fact that if each diagonal element of a matrix A is increased by a, the numbers λ comprising the field of values of the matrix will be increased by a.

Let L denote the best upper limit obtained for the absolute value of λ.

THEOREM 6. *Let a represent the absolute value of the determinant of A. Then*

$$a \leqq L^n.$$

This follows directly from the fact that $a = |\lambda_1\lambda_2 \cdots \lambda_n|$.

It is interesting to note that this result is frequently better than

the well known upper bound given by Hadamard: "If the elements of the matrix (a_{rs}) with $r, s \leqq n$ satisfy the condition $|a_{rs}| \leqq B$, then the following inequality holds:

$$|\det [a_{rs}]| \leqq B^n n^{n/2}."$$

Bibliography

1. Bendixson, *Sur les racines d'une équation fondamentale*, Acta Math. vol. 25 (1902) pp. 359–365.

2. Bromwich, *On the roots of the characteristic equation of a linear substitution*, Acta Math. vol. 30 (1906) pp. 295–304.

3. Browne, *The characteristic roots of a matrix*, Bull. Amer. Math. Soc. vol. 36 (1930) pp. 705–710.

4. Hausdorff, *Der Wertvorrat einer Bilinearform*, Math. Zeit. vol. 3 (1919) pp. 314–316.

5. Hirsch, *Sur les racines d'une équation fondamentale*, Acta Math. vol. 25 (1902) pp. 367–370.

6. Murnaghan, *On the field of values of a square matrix*, Proc. Nat. Acad. Sci. U.S.A. vol. 18 (1932) pp. 246–248.

7. Parker, *The characteristic roots of a matrix*, Duke Math. J. vol. 3 (1937) pp. 484–487.

8. Toeplitz, *Das algebraische Analogon zu einem Satze von Fejér*, Math. Zeit. vol. 2 (1918) pp. 187–197.

United States Military Academy